AFIT/GSO/ENS/92D-04

ANALYSIS OF WHOLE-SKY IMAGER DATA TO DETERMINE THE VALIDITY OF PCFLOS MODELS

THESIS

Richard William Boltz
Captain, USAF

George Richard Kather
Major, USA

AFIT/GSO/ENS/92D-04

Approved for public release; distribution unlimited

Thesis Approval

STUDENTS: Captain Richard W. Boltz
Major George R. Kather

CLASS: GSO-92D

THESIS TITLE: ANALYSIS OF WHOLE-SKY IMAGER DATA TO DETERMINE THE VALIDITY OF PCFLOS MODELS

DEFENSE DATE: November 23, 1992

COMMITTEE:	*NAME/TITLE/DEPARTMENT*	*SIGNATURE*
Reader	Brian W. Woodruff, Maj, USAF Associate Professor of Mathematics and Statistics	Brian W. Woodruff
Advisor	Thomas S. Kelso, LtCol, USAF Assistant Professor of Space Operations	Thomas Sean Kelso

AFIT/GSO/ENS/92D-04

ANALYSIS OF WHOLE-SKY IMAGER DATA TO DETERMINE THE VALIDITY OF PCFLOS MODELS

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

In Partial Fulfillment of the

Requirements for the Degree of

Master of Science (Space Operations)

Richard William Boltz, B.S.

Captain, USAF

George Richard Kather, B.S.

Major, USA

December, 1992

Acknowledgments

At this time we want to recognize and thank Jeffrey Yepez of Phillips Laboratories for providing the motivation and resources to allow us to complete this work. His expertise and assistance was instrumental in getting us started and in fostering a continuing institutional relationship between AFIT and Phillips Laboratories.

We also want to thank the members of our thesis committee, Lieutenant Colonel Thomas S. Kelso and Major Brian W. Woodruff. They kept us pointed in the right direction and provided a few clues when we were clue-less. We extend a special thanks to Lieutenant Colonel Kelso for adopting our thesis when our initial sponsorship dissolved—we are forever grateful.

Most importantly, we want to thank our wives, Cindy and Carol Ann, for their patience and personal sacrifice throughout our AFIT experience.

We were able to complete this study because we had initiative, self-discipline, and were willing to work long, hard hours. We want to dedicate this work to the men that taught us those qualities; namely, our fathers—Robert E. Boltz and Frank M. Kather.

One final note, a special thanks to Einstein and Newton.

Richard William Boltz

George Richard Kather

Table of Contents

List of Figures

List of Tables

Abstract

Models used to predict the probability of a cloud-free line-of-sight (PCFLOS) from the ground to space have existed since the 1960s. Unfortunately, an adequate data set has not been available to check the validity of these models until the deployment of the Whole-Sky Imager (WSI) system in 1989. Now that a three-year database has been collected from the WSI system, it is possible to validate, or refute, the existing models. This study investigates the most generally accepted models. Specifically, we investigate three questions: 1) Is the Lund and Shanklin PCFLOS model assumption of azimuthal independence valid; 2) Does the Lund and Shanklin sub-sampling of data via the use of a template adequately correlate to both the full image and the grid image; and, 3) Do the Lund and Shanklin and SRI model estimates correlate to the WSI observations. The primary contribution of this study is the development of a methodology which employs time series analysis techniques to evaluate and ultimately corroborate the assumption of azimuthal independence.

ANALYSIS OF WHOLE-SKY IMAGER DATA TO DETERMINE THE VALIDITY OF PCFLOS MODELS

I. Introduction

Background

In the past three years, the world has changed more than most people could ever have imagined. The Berlin Wall has fallen; Germany has unified, and perhaps most importantly, the Soviet Union has collapsed. This "outbreak of peace" has led to a major rethinking of our national defense strategy and subsequently, a reduction of our armed forces. General Colin L. Powell, Chairman of the Joint Chiefs of Staff, recently outlined the current reduction plan which amounts to roughly a 25 percent reduction in the overall size of our military forces (22).

In order to cut back personnel, yet still remain a dominant world power, the United States is becoming increasingly dependent on technologically-advanced "smart" systems. A great number of these systems are electro-optical in nature and, as a result, can be significantly affected by clouds. Both the "amount" or thickness of the clouds, and the variable occurrence of the cloud conditions contribute to the degree to which the performance of the electro-optical systems are degraded.

Most electro-optical systems require a cloud-free arc (CFARC) or cloud-free field-of-view (CFFOV) in order to complete their missions. For example, a ground-based anti-satellite (ASAT) laser would need a CFARC/CFFOV of a specific size in order to track and illuminate a target in space. Unfortunately, information on CFARCs is sparse because most cloud cover observations report only the fractional amount of clear sky; they do not describe the CFARC/CFFOV.

To maximize the performance of these electro-optical systems, a proper understanding of two key factors must be combined:

1. Effects that clouds have on specific regions of the electromagnetic spectrum.

2. A reliable evaluation of the spatial and temporal occurrences of cloud cover.

The effects that clouds and the atmosphere have on specific regions of the electromagnetic spectrum has been well characterized over the last 50 years (28). By contrast, only three substantial efforts have been made to characterize the spatial and temporal occurrences of cloud cover in the last 30 years. Furthermore, each of these efforts was limited by the lack of an adequate database from which to validate the analysis. In recognition of this lack of a cloud occurrence database, the Geophysics Directorate of Phillips Laboratory deployed a network of six whole-sky imager (WSI) systems to five locations in the continental United States (29:1). The WSI system is a visible-spectrum, computer-controlled, solid-state video camera that points upward with a fish-eye lens and has a nominal 180-degree field-of-view (29:1). At one-minute intervals, the camera records and archives a digital image of the sky. A cloud discrimination algorithm is then used to separate the images into five different intensity levels (missing, clear, thin-cloud, thick-cloud, and off-scale bright) (30:2). Now that this extensive database is available, it may be possible to validate existing models or develop a new theoretical model to predict the probability of a CFLOS and cloud-free intervals.

Summary of Current Knowledge

The earliest CFLOS studies were conducted independently by Lund and McCabe in 1965. The initial CFLOS estimates were derived from mean sunshine and mean total cloud cover observations (9). Because sunshine observations "see" through thin clouds, these estimates were only valid for the assumed case of exclusively opaque clouds. Accordingly, these initial models were limited in application and were not widely used. Since then, three significant models have evolved in the study of CFLOS: the Lund and Shanklin model (1973), the SRI model (1983), and the MOE model (1984). Each of these models are described in detail in the literature review (Chapter 2).

Two additional models have been developed as embellishments to the original CFLOS prediction models—the Boehm Sawtooth Wave Model (Boehm, 1986) and a procedure based on the Ornstein-Uhlbeck (O-U) class of Markov process presently being developed by Hering at the Marine Physical Laboratory of Scripps Institution of Oceanography. Both of these models are also discussed in the literature review for completeness. The study of duration, persistence, and joint occurrences and the associated CFARC and CFFOV estimates are embellishments of the more fundamental determination of the probability of a CFLOS. As many of the assumptions and approximations used in CFARC and CFFOV studies are based in part on PCFLOS models, it is a necessary first step to subject the basic PCFLOS models to close analysis against a suitable database.

Problem

The purpose of this thesis is to investigate models for use in predicting the probability of a cloud-free line-of-sight. Specifically, this study examines the currently accepted PCFLOS models which are a function of sky cover and zenith look angles—the Lund and Shanklin model and the SRI model.

Scope

The Whole-Sky Imager (WSI) system database will be used to validate candidate models. Use of the WSI system imposes the following limitations on the study:

- Cloud-free lines-of-sight will only be examined from a point on the earth's surface to a point in space. Cloud-free lines-of-sight between two points on the surface or between a surface station and aircraft will not be included in this study.
- Since the WSI system only collects data six hours before and after local apparent noon (LAN), the study is limited to daytime probability predictions.

- The five data collection stations for the WSI system are all located within the continental United States. Since weather patterns and cloud movements are influenced by the geographical region, the study only pertains to cloud-free lines-of-sight within the United States.

Assumptions

The data in the WSI database is assumed to be collected, processed, and stored as discussed in Appendix A. Accordingly, the stored data is assumed to accurately represent the presence or absence of clouds during the recorded periods. A discussion of the WSI system and data collection process is contained in Appendix A.

The data required to perform the statistical analyses will be made available by Phillips Laboratories.

Approach and Presentation

The ultimate goal of the study is to assess the correlation between the Lund and Shanklin and SRI PCFLOS estimates and the actual WSI system observations. Before this can be done, two underlying assumptions in the development of the Lund and Shanklin model must be addressed. Therefore, the general approach will be to divide the study into three research objectives and sequentially address each objective. The three research objectives to be examined are:

1. Azimuthal Independence. In developing their model, based on photos taken in Columbia, Missouri, Lund and Shanklin made the fundamental assumption that the probability of a CFLOS is independent of azimuth. This assumption has never been examined in detail, or tested, due to the lack of an independent database. Therefore, the first research objective is to rigorously test the azimuthal independence assumption.
2. Full Resolution Image versus Grid Subset. Lund and Shanklin used a subset of data from their photos to conduct their analysis, but never statistically demonstrated that the subset

of data accurately represents the full resolution image. The second research objective is to statistically evaluate the correlation between the WSI full resolution images (full images taken every 10 minutes), the sub-sampled images (grid images taken every minute), and the Lund and Shanklin template generated subset.

3. PCFLOS Model Estimates versus WSI Database Observations. The third research objective is to determine the correlation between the probability estimates as determined by the Lund and Shanklin and SRI models and the actual WSI database observations.

Each research objective will follow the same sequence of presentation and will include sections which describe the statistical model to be employed and the statistical procedures. The statistical model section will be described in terms of the theory, assumptions, limitations, and applicability of the model.

The methodology for each research objective will be presented in the first three subsections of Chapter 3.

In Research Objective Three, we will assess the correlation between the theoretical model estimates and the WSI database observations by graphically comparing the PCFLOS estimates and evaluating the differences at each elevation angle. The PCFLOS theoretical estimates will be generated by varying the azimuth, elevation, and total sky cover variables as follows:

- Elevation. The elevation will be varied in 10-degree increments from 0 to 70 degrees from zenith.
- Azimuth. The azimuth parameter will be set to 0, 90, 180, and 270 degrees.[1]
- Total Sky Cover. The total sky cover variable will range from 0 percent sky cover to 100 percent sky cover in 10-percent increments.

[1] These azimuths are consistent with the azimuthal directions evaluated in the WSI database.

Chapter 3 closes with the generation of PCFLOS estimates from the WSI database. The WSI database will be manipulated to extract a usable subset of data for comparison against the theoretical estimates. WSI images will be sorted according to the total sky cover percentage (in tenths). Statistical analyses will then be performed on the observations recorded at the azimuths and elevations described above.

The resultant PCFLOS estimates from the theoretical models and WSI database will be displayed for comparison in Chapter 4.

Finally, the accuracy, applicability, and overall validity of the theoretical models (along with recommendations for improvement and further research) will be discussed in Chapter 5.

II. Literature Review

The Lund and Shanklin Model

In the early 1970s, Lund and Shanklin determined the relative frequencies of cloud-free lines of sight (CFLOS) at specified elevation angles by examining whole-sky photographs taken at hourly intervals during the summer months from 1966 through 1969.

The photographs used in this study were taken from the United States National Weather Service (NWS) observing site, located at Columbia, Missouri, using a 180-degree (fish-eye) lens and infrared film. These high-contrast, whole-sky exposures were made at five minutes before the hour, concurrent with the NWS observations. The photographs were collected between sunrise and sunset from 1 March 1966 to 28 February 1969 and from 1 June 1969 to 31 August 1969 (16:774).

Lund and Shanklin placed a clear plastic template over each of the photographs to pinpoint the exact locations of the lines-of-sight on the prints (16:775). The template contained 33 small circles whose centers represent the 33 lines-of-sight at azimuths of 0°, 90°,180°,and 270° and elevation angles of 10° to 90°, in 10° increments (see Figure 2.1) (16:776).

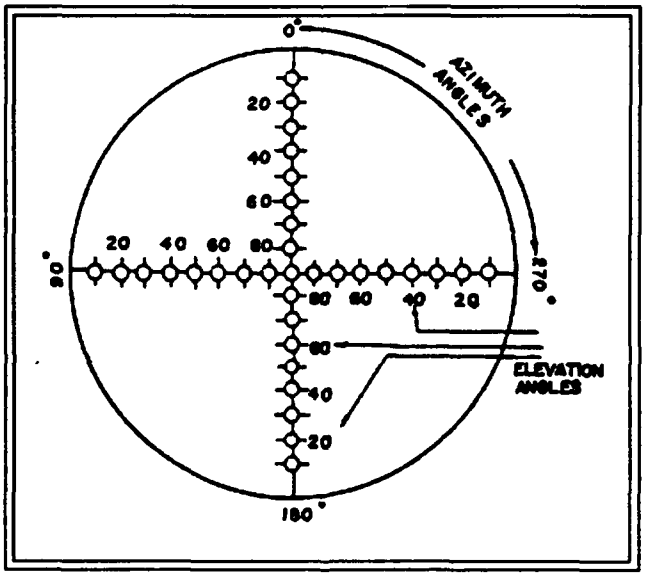

Figure 2.1. The Lund and Shanklin template.

Lund and Shanklin found the relative frequency of a CFLOS for each elevation angle as a function of the tenths of cloudiness as reported by the NWS (16:777). Their results can be found in Table 2.1. Note that in Lund and Shanklin's results, the sample size is four times the number of photographs examined. This is because the observations along each of the four cardinal directions were counted as separate sets of data. This quadruple counting implies that Lund and Shanklin assumed azimuthal independence.

From this data set, Lund and Shanklin plotted the probability of a cloud-free line of sight versus elevation angle for each tenth of total sky cover (see Figure 2.2). These curves were then "subjectively smoothed" (16:781). The smoothed curves are depicted in Figure 2.3.

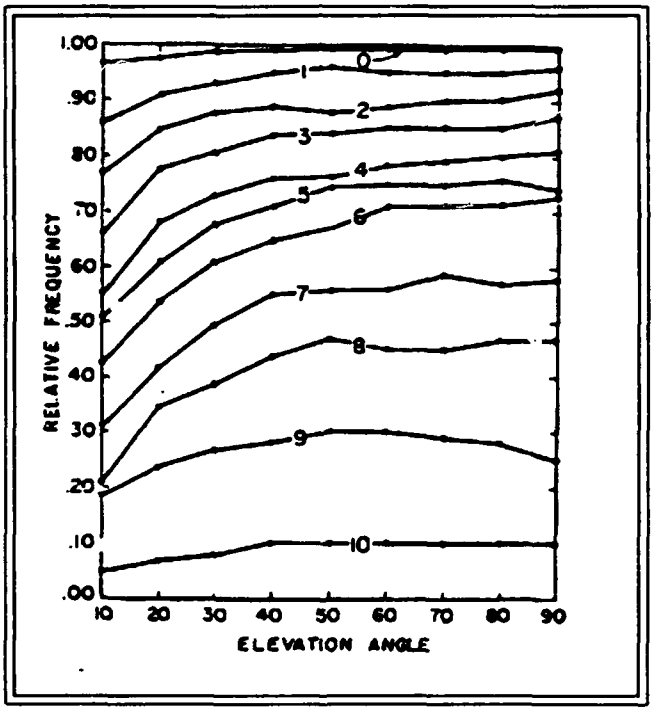

Figure 2.2. Relative frequencies of cloud-free lines-of-sight (CFLOS) as a function of elevation angle and observed total sky cover, in tenths.

Depending on the type of clouds reported by the NWS, the data from the photographs were divided into six cloud form categories (see Table 2.2) (17:30). Tables similar to Table 2.3 were prepared for each of the cloud form categories.

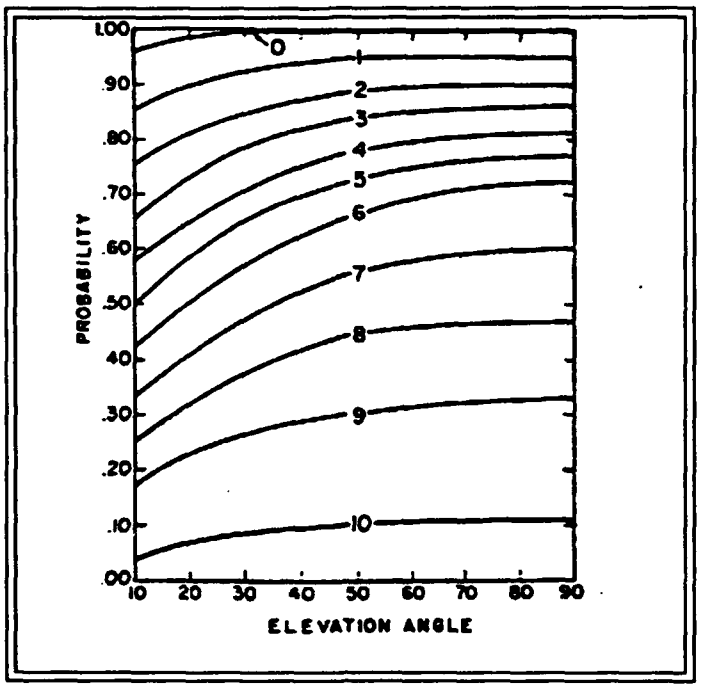

Figure 2.3. Estimated probabilities of CFLOS as a function of elevation angle and National Weather Service observed total sky cover.

Lund and Shanklin developed two methods for estimating probabilities of CFLOS (PCFLOS). Method A was to be used for geographic locations with cloud type frequency distributions similar to the distributions observed at Columbia, Missouri. Method B was to be used for all other locations. However, according to Donald Grantham of Phillips Laboratories, the cloud type frequency distribution only contributes second-order effects to the PCFLOS model (10:1). Therefore, only Method A will be discussed in this review.

Lund and Shanklin go on to state that the probabilities of CFLOS can be estimated with the following formula:

$$ {}_{\alpha}\hat{P}_1^A = {}_{\alpha}C_{ss}K_1 $$

where ${}_{\alpha}\hat{P}_1^A$ is a column vector of α rows, one for each elevation angle considered; ${}_{\alpha}C_s$ is a matrix of α rows and s columns, one row for each elevation angle, one column for each sky cover

Table 2.1. Relative frequencies of cloud-free lines-of-sight as a function of elevation angle and National Weather Service observed total sky cover.

Total Sky Cover (tenths)	*Elevation Angle*									*Sample Size*	*Relative Frequency*
	10°	20°	30°	40°	50°	60°	70°	80°	90°		
0	0.97	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	2044	0.153
1	0.86	0.91	0.93	0.95	0.96	0.95	0.95	0.95	0.96	832	0.062
2	0.77	0.85	0.88	0.89	0.88	0.89	0.90	0.90	0.92	900	0.067
3	0.66	0.78	0.81	0.84	0.84	0.85	0.85	0.85	0.87	876	0.066
4	0.55	0.68	0.73	0.76	0.76	0.78	0.79	0.80	0.81	700	0.052
5	0.51	0.61	0.68	0.71	0.75	0.75	0.75	0.76	0.74	588	0.044
6	0.42	0.54	0.61	0.65	0.67	0.71	0.71	0.71	0.73	848	0.064
7	0.31	0.42	0.50	0.55	0.56	0.56	0.59	0.57	0.58	916	0.069
8	0.21	0.35	0.39	0.44	0.47	0.45	0.45	0.47	0.47	1104	0.083
9	0.19	0.24	0.27	0.28	0.30	0.30	0.29	0.28	0.25	1008	0.076
10	0.05	0.07	0.08	0.10	0.10	0.10	0.10	0.10	0.10	3532	0.265
Average	0.44	0.51	0.54	0.56	0.56	0.57	0.57	0.57	0.57	13,348	

category; and ${}_{s}K_1$ is a column vector of s rows. The $\hat{P}$ values are estimates of the probabilities of CFLOS through the atmosphere, the C values are probabilities of CFLOS at angles α given k tenths of cloudiness (see Table 2.3[1]), and the K values are probabilities of each k tenths of cloudiness (17:32).

As shown in the formula, the probability of a CFLOS is strictly a function of elevation angle and total sky cover. Lund and Shanklin's work assumes that the cloud frequency distribution is independent of the azimuth. This assumption has never been validated. Also, the Lund and Shanklin results have never been independently tested because an adequate, independent database was previously not available.

[1]The data in this table was extracted from curves which resulted from the smoothing of the data in Table 2.1.

Table 2.2. Cloud-form categories.

Category	*Form*	*Cloud type*
1	Cirriform	Cirrocumulus
		Cirrostratus
		Cirrus
2	Middle	Altocumulus
		Altocumulus castellanus
		Altostratus
3	Cumuliform	Cumulonimbus
		Cumulonimbus mammatus
		Cumulus
		Fractocumulus
4	Stratiform	Fractostratus
		Nimbostratus
		Stratocumulus
		Stratus
5	Mixed	Mixtures of more than one form
6	None	No clouds of any type reported

Table 2.3. Probabilities of cloud-free lines-of-sight as a function of elevation angle and observed total sky cover, when all cloud types are included in the data sample.

		Total sky cover (tenths)										
		0	1	2	3	4	5	6	7	8	9	10
	90	1.00	0.97	0.92	0.87	0.81	0.77	0.70	0.62	0.48	0.31	0.08
	80	0.99	0.97	0.92	0.87	0.81	0.77	0.69	0.61	0.47	0.31	0.08
	70	0.99	0.97	0.91	0.86	0.80	0.76	0.68	0.61	0.47	0.30	0.08
Elevation	60	0.99	0.96	0.90	0.85	0.80	0.75	0.66	0.60	0.46	0.29	0.08
angle	50	0.99	0.96	0.90	0.85	0.78	0.73	0.64	0.58	0.45	0.29	0.08
(degrees)	40	0.99	0.95	0.88	0.83	0.76	0.71	0.62	0.55	0.42	0.27	0.07
	30	0.98	0.93	0.86	0.80	0.73	0.66	0.57	0.50	0.38	0.21	0.05
	20	0.98	0.90	0.83	0.75	0.67	0.59	0.50	0.42	0.33	0.21	0.05
	10	0.97	0.86	0.76	0.65	0.55	0.47	0.39	0.32	0.24	0.16	0.03

The Allen and Malick (SRI) model

Using the Lund and Shanklin data as "truth," Allen and Malick developed a model that fits (quite well) the ${}_{\alpha}\hat{P}_1$ data as shown in Table 2.1 (18:1).

They argued that if q_d is the probability of a straight line passing through a cloud over a distance d of a homogeneous volume, then the probability of intercepting a cloud over a path of n independent segments, each d in length is:

$$Q = 1 - (1 - q_d)^n$$

from which it follows,

$$PCFLOS = 1 - Q = (1 - q_d)^{R/d} = e^{(-q_d R/d)}$$

(where both R and d are functions of the elevation angle of the line projected through the atmosphere) (18:1).

Assuming an average height-to-width ratio, b (for a cubical cloud form), Allen and Malick showed that the projected area of the cloud (and therefore q_d) at an elevation angle α is proportional to $\sin\alpha + b\cos\alpha$ (18:1). They combined this with the fact that the path length increases with $1/\sin\alpha$ to come up with the relationship:

$$PCFLOS = P_n^{(1 + \frac{b}{\tan\alpha})}$$

where,

P_n is the probability of CFLOS at zenith.

Allen and Malick empirically derived best-fit values for P_n and b and subsequently generated a ${}_{\alpha}\hat{P}_1$ matrix which was consistent with the Lund and Shanklin data (18:1). Upon integration over 2π steradians and setting equal to 1 - s, Allen and Malick were able to generate PCFLOS values.[2]

The values for P_n and b are as follows:

$$b = 0.55 - \frac{s}{2}$$

$$P_n = 1 - s\frac{(1+3s)}{4}.$$

Putting them all together we obtain,

$$PCFLOS = (1 - \frac{s(1+3s)}{4})(1 + (0.55 - \tfrac{s}{2})\tan\alpha)$$

The major shortcoming of this model is that it uses the results of the Lund and Shanklin model to empirically derive the model parameters. Accordingly, the accuracy of the model is limited to the inherent accuracy of the Lund and Shanklin model.

[2]*Since s is the percentage of sky cover in tenths, $1 - s$ is the percentage of clear sky in tenths. Originally, the* ${}_{\alpha}\hat{P}_1$ data did not integrate to 1 - s, so Allen and Malick adjusted the data by setting it equal to $1 - s$.

The MOE (Estonian) Model (8)

Researchers from the Institute of Astrophysics and Atmospheric Physics of the Estonian Academy developed a theoretical model which describes the statistical structure of cumulus fields. The authors defined an indicator function $n(\theta, \phi, x, y, t)$, which when applying the theory of random processes to this function, allowed them to empirically derive various statistical characteristics of cumulus fields. The most pertinent relationship is an empirical formula that relates the probability of cloud cover at a given elevation (off-zenith), $n(\theta)$, to the mean cloud cover at zenith, n(0), the elevation angle off-zenith, θ, and a coefficient, b, which is a function of the vertical depth of the clouds:

$$n(\theta) = 1 - [1 - n(0)]e^{(-n(0)b(\sec(\theta) - 1))}$$

This empirical relationship was derived from an unknown number of sky photos taken on land in the Estonian SSR and in the Atlantic (27N, 25W). In deriving the relationship, the authors assumed statistical isotropy of the cloud field and assumed the variability of the cloud field as a normal random process (8).

The MOE model closely parallels the PCFLOS initiatives, but relates the PCFLOS as a function of elevation angle and sky cover at the zenith (versus total sky cover). Due to the fact that the original photos were taken over both land and water and no information is available about the photos (quantities and proportion land or water) this model is considered inappropriate for analysis and comparison with the Lund and Shanklin and SRI models. Consequently, the MOE model will not be evaluated in this study.

The Boehm Sawtooth Wave Model (2)

In 1986, Boehm, and others, developed a comprehensive statistical model that provides the *duration* of CFLOS from ground sites to geostationary satellites. This model uses the multidimensional Boehm Sawtooth Wave Model to establish climatic probabilities through repetitive simulations of sky cover distributions. As emphasized above, the focus of this model was the estimation of how long a CFLOS would exist. Therefore, this model is an extension of the more basic PCFLOS estimation problem and will be excluded from further discussion in this paper.

The (O-U) Markov Model (11)

Gringorten (1966, 1968, 1972) developed a modeling procedure based on the Ornstrein-Uhlbeck (O-U) class of Markov process to estimate the *joint occurrence and duration* of a variety of weather events. Hering of the Marine Physical Lab developed an extension of the O-U Markov model to estimate the *joint occurrence and persistence* probabilities of CFLOS (13). Hering's work focuses on calculating probability estimates of the *duration* of cloud-free lines-of-sight from one or multiple ground sites. Hering's application of the (O-U) Markov model to the temporal domain (persistence) is, in fact, an area of study unique unto itself. Since Hering's efforts do not include PCFLOS estimation directly, this model will not be discussed further in this study.

III. Methodology

As stated in Chapter 1, the purpose of this study is to investigate models for use in predicting the probability of a cloud-free line-of-sight. For the reasons cited in Chapter 2, we have narrowed our attention to the Lund and Shanklin and SRI models and focused our investigation on three research objectives—azimuthal independence, subsampling, and model estimates. The purpose of this chapter is to explain the procedures used to analyze each of the three research objectives. This chapter is divided into three sections—one for each research objective.

Azimuthal Independence Tests

We opted to use three separate approaches to examine the validity of Lund and Shanklin's assumption of azimuthal independence.

- A χ^2 test of proportionality.
- Trend analysis.
- Time series analysis.

The first two tests are tests of proportionality and the third test is an application of time-series analysis techniques. Tests of proportionality were selected because of the form of our data (ordinal values–clear, thin-cloud, and thick-cloud). The methodology for each of the three approaches are described in detail below.

χ^2 Test of Proportionality.

1. Statistical Model and Theory. For our application, we are interested in determining if the probability (or proportion) of the observations being in a clear, thin-cloud, or thick-cloud category is the same or different among populations drawn from the north, east, south, and west directions. The rows of the contingency table represent the four cardinal azimuths and

the columns represent the observations (clear, thin-cloud, thick-cloud) as depicted in Table 3.1. Each cell in the table represents the number of observations belonging to both the row and column categories.

Once the data is sorted in the contingency table according to the two criteria, a *test statistic*, T, can be derived from:

$$T = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where,

$E_{ij} = \frac{n_i C_j}{N}$, ($n_i$ is the number of observations along the i^{th} azimuth, and N is the total number of observations (see Table 3.1).)

O_{ij} = the *observed* number in cell ij

E_{ij} = the *expected* number of observations in cell ij, if H_o is really true (6:151).

Table 3.1. Data arrangement for a $r \times c$ contingency table.

Azimuth	*Observation*			
	Clear	*Thick-Cloud*	*Thin-Cloud*	*Total*
East	O_{11}	O_{12}	O_{13}	n_1
North	O_{21}	O_{22}	O_{23}	n_2
South	O_{31}	O_{32}	O_{33}	n_3
West	O_{41}	O_{42}	O_{43}	n_4
Total	C_1	C_2	C_3	N

For large sample distributions (where the E_{ij} quantities are larger than five) the χ^2 statistic can be used to approximate the critical region. For a given level of significance, α, the rejection region is where T exceeds $\chi_{1-\alpha,\nu}$ (where ν represents the degrees of freedom) (6:152). Since all of our expected values are large, the χ^2 test of proportionality is an appropriate test.

2. Hypothesis. The null and alternate hypotheses are:

H_o: The proportion of observations (clear, thin-cloud, thick-cloud) are the same.

H_a: The proportion of observations (clear, thin-cloud, thick-cloud) are not the same.

We chose to examine the χ^2 statistic at a 0.05 level of significance. Therefore, our rejection criterion was to reject H_o if χ^2 >12.59, the value of $\chi^2_{0.05}$ for $(3-1)(4-1) = 6$ degrees of freedom.

3. Model Assumptions. Two assumptions must be satisfied to use a χ^2 test of independence or proportionality.

   - The sample of N observations is a random sample. This prerequisite is satisfied by our sampling methodology. We considered the year of available data from the four cardinal directions as the total population. In sampling this total population, we decided to use all the data points. Thus, each sample had an equal opportunity to be drawn.
   - Each observation may be classified into exactly one of r categories according to one criterion, and into exactly one of c categories according to a second criterion. This assumption is satisfied because each of our data points is discretely categorized by observation (clear, thin-cloud, thick-cloud) and by azimuth (north, east, south, and west).

4. Model Limitations. The data used for the observation in a $r \times c$ contingency table must be from a nominal (or higher) scale of measurement. As our data consists of ordinal data, this limitation does not preclude our use of the χ^2 test of proportionality with a $r \times c$ contingency table.

5. Model Applicability. The $r \times c$ contingency table is a statistical procedure used to analyze two or more samples drawn from different populations to see if the populations have the same proportion of elements in a certain category (6:141). In our context, the different populations are the different azimuths and the categories are the observations.

6. Statistical Procedures. Contingency tables, as described above, were set up using azimuth and observations as the categories for r and c, respectively. Data points were accumulated

in each cell according to two time intervals—monthly and yearly (from February 1989 to January 1990). The monthly intervals were evaluated to ensure that the fidelity of the data did not "average out" over the extended, yearly period. The test design was to accumulate observations at elevations of 10, 30, 50, and 70 degrees off-zenith at the monthly and yearly frequencies. For example, the contingency table for the cumulative time period at 10-degree off-zenith is depicted in Figure 3.1 with each cell containing the actually observed count. Note that the χ^2 statistic value for this configuration was 573.227. The χ^2 statistics for each of the monthly and cumulative configurations have been tabulated and are presented in Chapter IV.

	Frequency| Percent | Row Pct | Col Pct	OBSERVATION |CLEAR	|THICK	|THIN	Total
AZIMUTH	EAST	124507	83567	23213	231287
		13.66	9.17	2.55	25.38
		53.83	36.13	10.04	
		25.70	24.45	27.32	
	NORTH	120658	89357	21164	231179
		13.24	9.81	2.32	25.37
		52.19	38.65	9.15	
		24.91	26.14	24.91	
	SOUTH	115018	83795	19040	217853
		12.62	9.20	2.09	23.91
		52.80	38.46	8.74	
		23.74	24.51	22.41	
	WEST	124227	85101	21560	230888
		13.63	9.34	2.37	25.34
		53.80	36.86	9.34	
		25.65	24.90	25.37	
	Total	484410	341820	84977	911207
		53.16	37.51	9.33	100.00

STATISTICS FOR TABLE OF AZIMUTH BY OBSERV

Statistic	DF	Value	Prob
Chi-Square	6	573.227	0.000

Sample Size = 911207

Figure 3.1. Contingency table for the cumulative data at Columbia, Missouri from February 1989 to January 1990 (elevation = 10 degrees off-zenith).

Trend Analysis.

1. Statistical Model and Theory. In their PCFLOS model, Lund and Shanklin considered the sky-dome isentropic, and therefore, not a function of azimuth. To test this assumption, we started our investigation by plotting the data to identify trends, cycles, and correlations. Because trend identification is the only goal of this test, no statistical models or test statistics are employed.

2. Hypothesis. If Lund and Shanklin's assumption of azimuthal independence is true, we would expect the rate of change of an observation along each cardinal axis to be the same.[1] For example, if the observation at zenith is clear, then the observation at 10° off-zenith will also be clear with some measure of correlation. The same would be true for observations taken at 10° when compared to its first-order neighbors (zenith and 20° off-zenith). If the sky-dome is truly isentropic, the rates of change of an observation should be the same in every direction. Therefore, our null and alternative hypotheses are:

   H_o: The rate of change of the observation is the same in each cardinal direction.

   H_a: The rate of change of the observation is not the same in each cardinal direction.

3. Model Assumptions. No test statistics or models were applied to this test.

4. Model Limitations. Since we do not use a test statistic to rigorously measure the degree of "sameness" along each direction, we are inherently limited in what we can discern about the "sameness" along each cardinal axis. But, as stated above, we are simply interested in observing trends as an *indicator* of the degree of isentropy.

5. Model Applicability. This approach is very limited in its applicability. The results of this test are strictly indicators to help focus our efforts in follow-on testing.

[1] The "same" to a specified degree of statistical significance.

6. Statistical Procedures. A working data set was developed from the Columbia, Missouri archived data. The working data set stored the observations along each of the cardinal directions at elevation angles of 10° to 70° off-zenith. Observations were at one-minute intervals from February 1989 to January 1990. A spreadsheet was used to sum the data by month and produce plots of the percentage of clear days (see Figure 3.2).

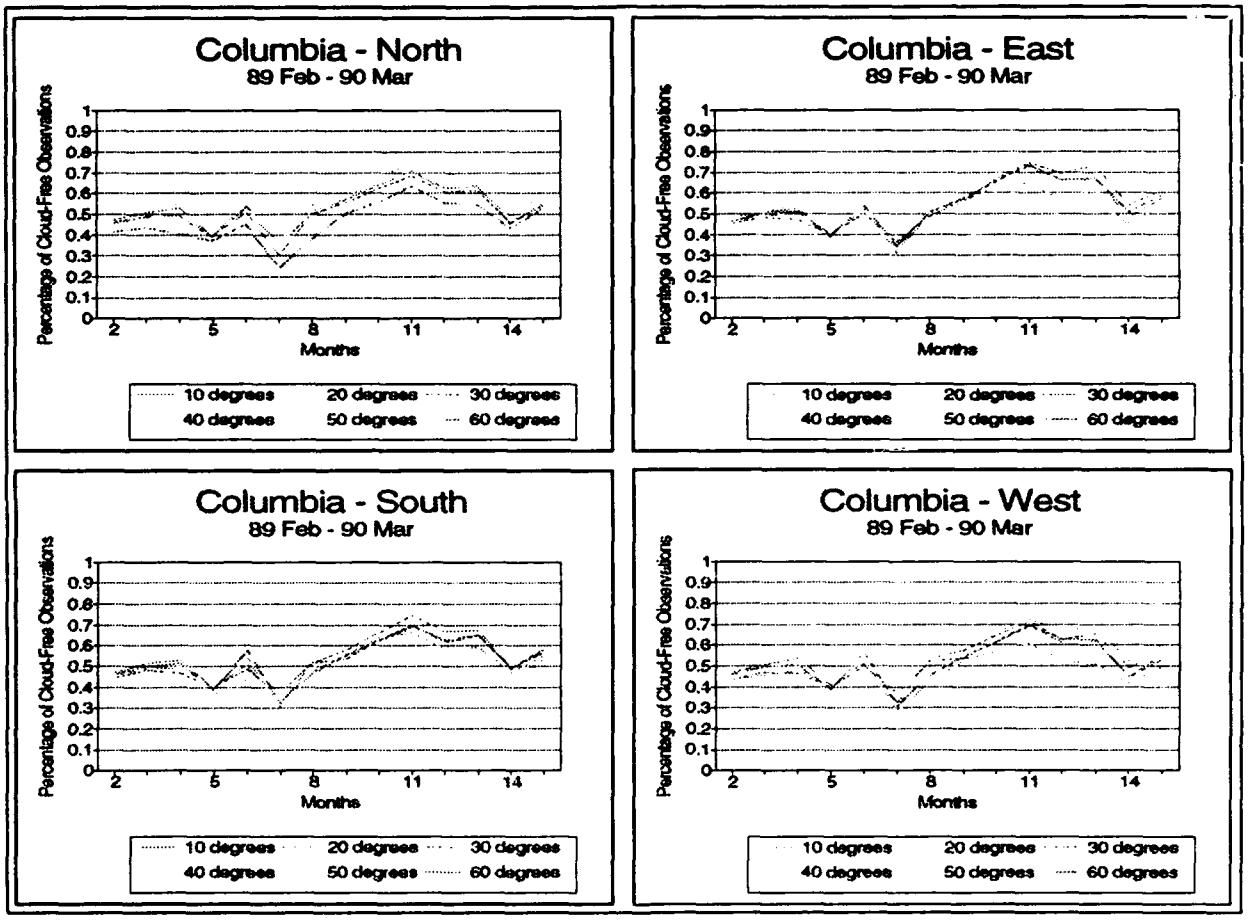

Figure 3.2. Plots of the percentage of clear days for Columbia, Missouri from February 1989 to January 1990.

From these plots we noted that a strong correlation appears to exist between the observations at different elevation angles along each axis. To examine this further, we used the spreadsheet to determine whether the neighboring observation is the same for lags of 1 through 7. For example, the seven data points that comprise the lag = 1 column are produced by taking the

10° observation and comparing it to the 20° observation to see if they are the same. If they are the same, a value of one is assigned. If they are not the same, a value of zero is assigned. The same approach is applied taking each elevation from 20° to 70° as the *base* observation for comparison. Lags 2 through 7 are produced in a similar fashion with a diminishing number of data points due to the larger neighbor distances.[2] The "ones and zeros" are then summed and the percentage of "same" observations are plotted versus the lags for each direction (see Appendix E). The data points for each lag are then plotted about their mean with a one standard deviation confidence interval (also displayed in Appendix E).

Time Series Analysis.

1. Statistical Model and Theory. As discussed by Box and Jenkins (1976) the purpose of time series analysis is to model the dependence between observations in a time series.[3] For example, a time series of observations (such as stock market values recorded over a calendar year) can be modeled with appropriate autoregressive (AR) and moving average (MA) terms. Once the data is fitted to an appropriate model, the model can then be used to forecast into the future. Building an appropriate model is an iterative process consisting of three parts—model identification, model estimation, and diagnostic checking (3:171).

    - Model Identification. Model identification is accomplished primarily by examining plots of the autocorrelation function (ACF) and partial autocorrelation function (PACF). The ACF reveals how the correlation between any two values of a series change as their separation changes (3:30). A duality exists between the ACF and PACF which is useful in identifying models (3:72). An AR = 1 term is indicated if the plot of the ACF tails off in damped sine waves or exponentials and the PACF plot cuts off after the first p terms. From duality, a MA = 1 term is indicated if the plot of the ACF cuts off after the

[2] Lag 2 compares second-order neighbors, Lag 3 compares third-order neighbors. . .
[3] For a detailed discussion of time series analysis, the reader is referred to Box and Jenkins (1976).

first q terms and the PACF plot tails off in damped sine waves or exponentials (3:79). As noted by Box and Jenkins, we followed the methodology that it is only necessary to difference the original data twice (at most) in most practical applications and that it is sufficient to only examine the ACFs and PACFs for the first 20 lags (3:175).

- Model Estimation. Statgraphics®software routines were used to estimate the models and determine the coefficients of the significant terms. Statgraphics®employs the ACF, PACF, and model estimation techniques developed by Box and Jenkins (1976).[4]

- Diagnostic Checking. Once a model has been estimated, the model must be checked to ensure it adequately represents the data set, yet does not overestimate the data set. As noted by Brockwell (1991) the more terms added into the model, the more closely the model will estimate the data. But, on closer inspection, this results in "over-fitting" the data (5:287). In the extreme, over-fitting produces gross errors in future predictions. So, the objective of diagnostic checking is two-fold: to ensure the model adequately describes the original data; and, to minimize the terms of the model.[5] To determine the best model, we examined the ACFs, PACFs, normal probability plots, and periodograms of the residuals and performed five statistical tests. The statistical tests are elaborated upon in the Statistical Procedures section below.

To investigate the dependence of observations on azimuth, we employed time-series analysis techniques to develop an ARIMA[6] model for each direction at the Columbia and Kirtland sites. Then, the models can be compared and analyzed to determine similarities, if any.

2. Hypothesis. Any stationary time-series can be modeled with appropriate AR and MA terms. ARIMA processes can be defined by the equation

[4] A detailed explanation of the formulas employed can be found in Appendix F.
[5] This concept is referred to as parsimony.
[6] ARIMA is the acronym for AutoRegressive, Integrated, Moving Average.

$$\phi(B)(1-B)^d z_t = \theta_0 + \theta(b)a_t$$

where,

$\phi(B)$ and $\theta(B)$ are backshift operators of degree p and q, respectively.

For our purpose of determining isentropy, we are interested in the *spatial* relationship (dependence of observations) from zenith out to the horizon. Therefore, we developed a "pseudo" time series by concatenating the observations in 10° increments from zenith to 70° along each cardinal direction. If the sky-dome is truly isentropic, the model that describes the "pseudo" time series in each direction should be the same model and the variance in the coefficients of each term in the model should be very small (statistically).

3. Model Assumptions. The dataset must be stationary. This assumption was satisfied by differencing the raw data. As discussed below in the statistical procedures, the data was differenced to take out any seasonal effects. Mathematical definitions of covariance stationarity and "strict" stationarity are discussed by Brockwell (5:12). For our purposes, Brockwell's intuitive description is sufficient. Brockwell states that if a time series is stationary, two equal-length time intervals of the series should exhibit similar statistical characteristics.

4. Statistical Procedures. In the "raw data" field displayed in Figure 3.3, the percentage of clear observations at elevation angles in 10° increments were linked together to form a *spatial* analog to a *time*-series.[7] This time-series was subsequently differenced twice to obtain a stationary data set (see Figure 3.3). From the plots, we noted that the second differencing resulted in larger variations and concluded that only a first differencing was necessary to obtain a stationary data set. A sample of the original series and differenced data is depicted in Figure 3.3. A complete set of differenced data plots is included in Appendix F.

[7] The seven (10° to 70° in 10° increments) frequency plots were concatenated together to form a "pseudo" time-series. This was based on the strong correlation between observations at different elevation angles along each cardinal axis as demonstrated in our Trend Analysis section.

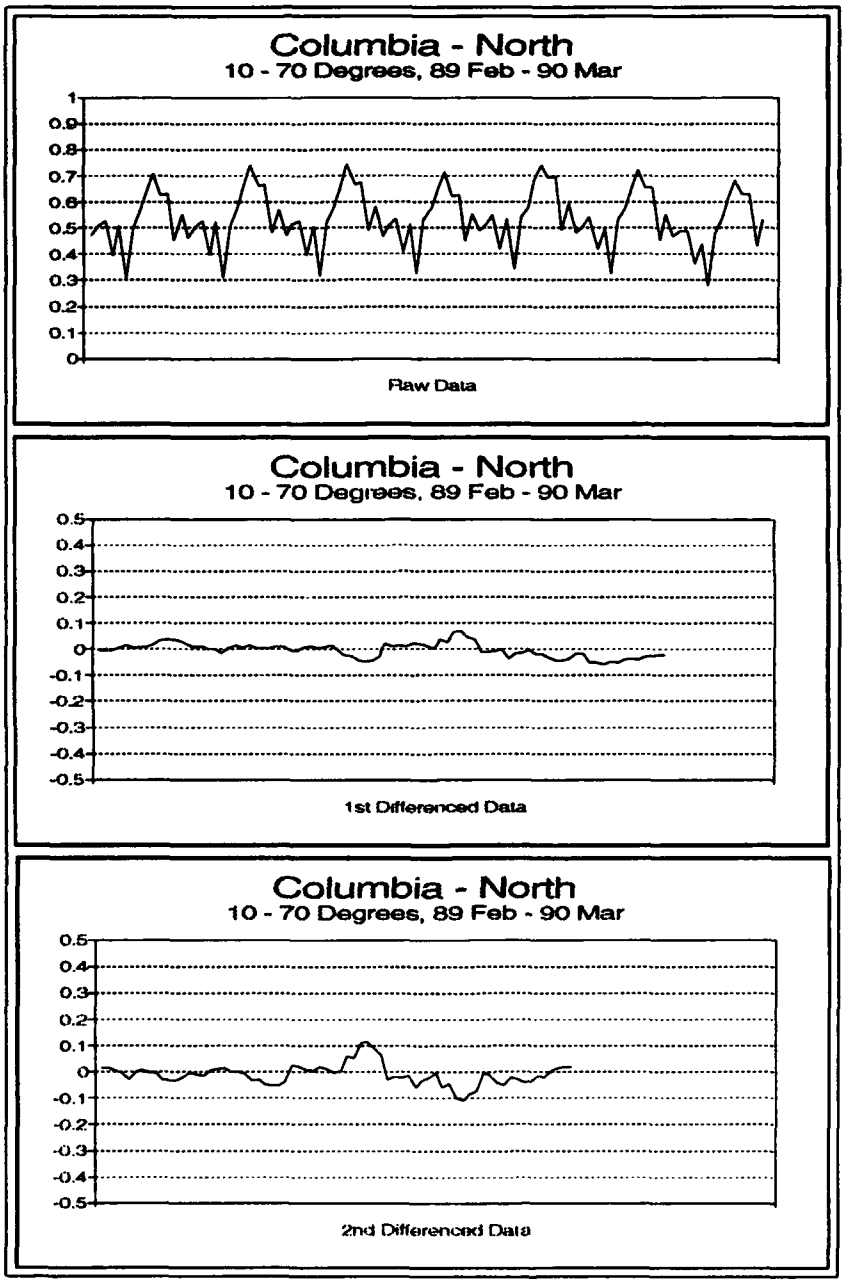

Figure 3.3. Raw and differenced data for the north azimuth at Columbia, Missouri.

Table 3.2. ARIMA models estimated for each cardinal direction.

Test	*ARIMA Model Parameters*		
	p	*d*	*q*
a	0	0	1
b	0	0	2
c	1	0	0
d	1	0	1
e	1	0	2
f	2	0	0
g	2	0	1
h	2	0	2

Once a stationary data set was obtained, Statgraphics®routines were employed to calculate and plot the ACFs, PACFs, and periodograms for the original (differenced) series. We chose to only examine the first and second order ARIMA models shown in Table 3.2.[8]

The ACFs, PACFs, and periodograms for the original (differenced) series were then examined to determine a subset of candidate models. The ACF and PACF plots are used to indicate the p and q terms required in the model. For an AR process, the ACF tails off and the PACF cuts off after lag q. For a MA process, the PACF tails off and the ACF cuts off after lag p. If both the ACF and PACF tail off, a mixed ARMA model is suggested (3:175). The periodogram shows the frequency spectrum of the series. High frequency responses indicate the model has a periodicity. For residual plots, the periodogram should be uniformly distributed if the residuals are truly "white" noise. Samples of these plots are depicted in Figures 3.4 and 3.5, respectively. A complete display of the plots for all four cardinal directions at both sites is included in Appendix F.

The candidate models were subsequently tested to determine the "best" model by employing five tests on the residuals:[9]

[8]The original data ACFs and PACFs did not indicate the presence of any higher significant terms. So, higher-termed models (for example, $ARIMA(3,0,0)$ were only randomly investigated. The higher-termed models characteristically injected more structure into the lower lags of the corresponding ACFs and PACFs and were rejected as feasible models.

[9]The code used to conduct these tests was written by Dr. T.S. Kelso.

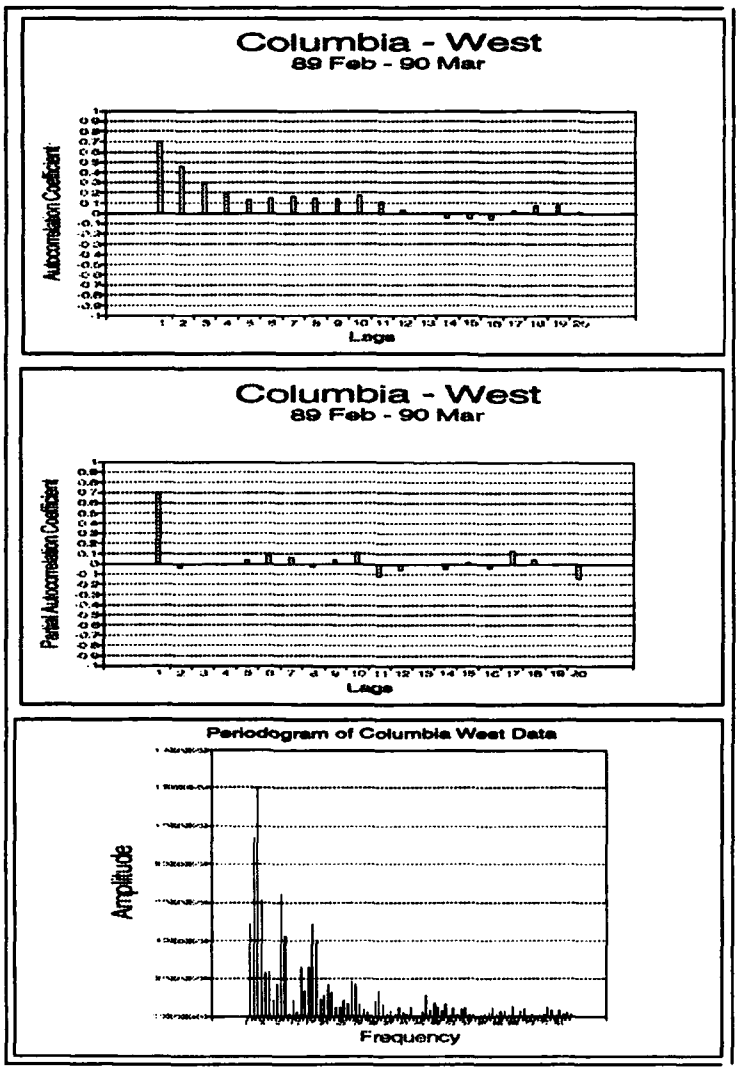

Figure 3.4. A sample of the ACF, PACF, and periodograms for the original data.

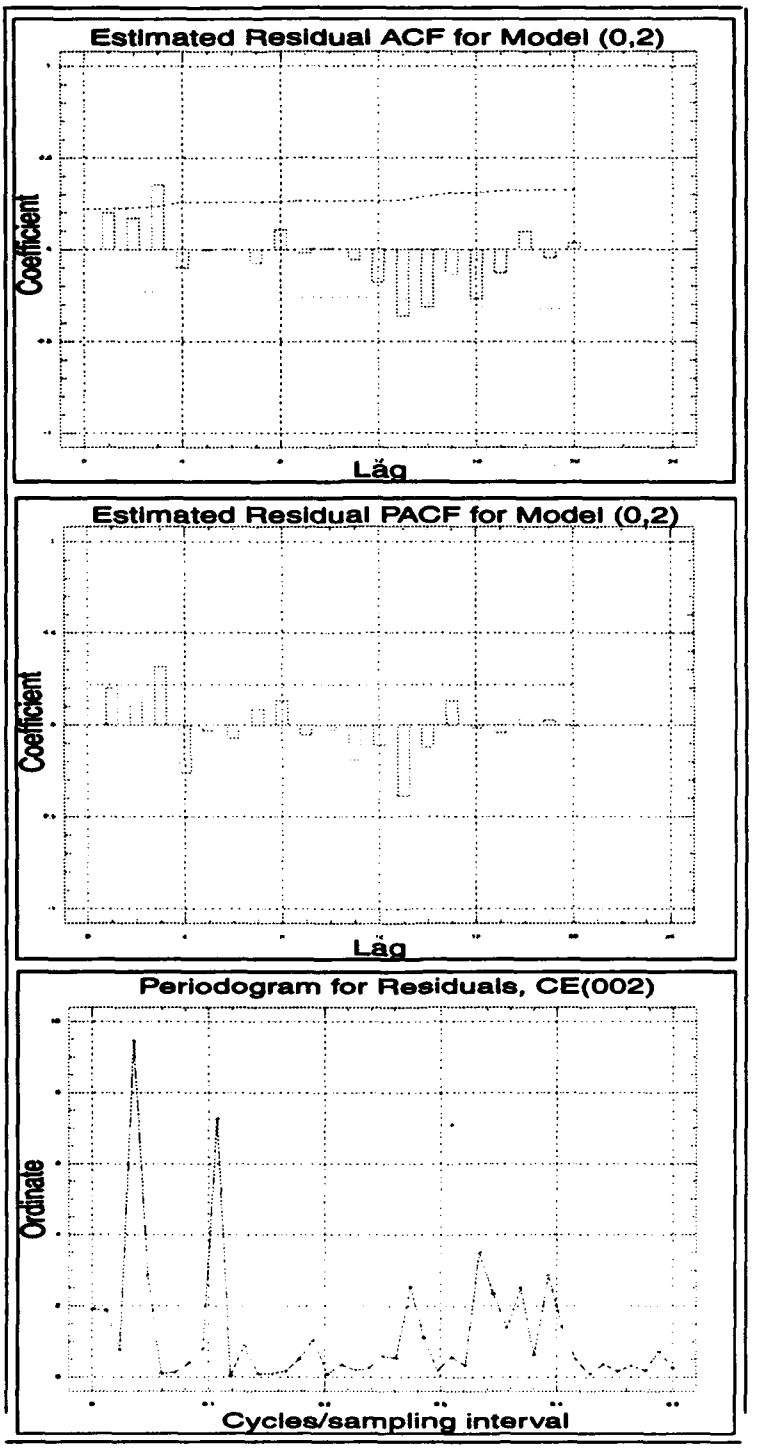

Figure 3.5. A sample of the ACF, PACF, and periodograms for the residual data.

- Runs Up and Down Test.
- Runs Above and Below the Mean Test.
- Frequency Test.
- Portmanteau Test.
- Akaike Information Criterion (AIC).

Runs Up and Down Test

(a) Statistical Theory. A *run* in a sequence of symbols is a group of consecutive symbols of one kind preceded and followed by different symbols. For our application, the residual values are the sequence of interest. Each consecutive value in the residual sequence, r_i, can be evaluated to see if it is greater than or less than the preceding value (15:498). A binary sequence, S, can then be constructed assigning the ith term in S a value of *zero* if $r_i < r_{i+1}$ and is equal to *one* if $r_i > r_{i+1}$. A run of length k is formed by a subsequence of consecutive zeroes bracketed by ones on each end. Likewise, a run of ones of length k is formed by a subsequence of consecutive ones bracketed by zeroes. The number of runs of length k can then be counted for the sequence S. The runs up and down test compares the actual (counted) number of runs encountered to the theoretical "expected" values (21:60). The expected values based on a "truly" random sample are:

for total runs,

$$\frac{(2N-1)}{3}$$

for runs of length 1,

$$\frac{(5N+1)}{12}$$

for runs of length 2,

$$\frac{(11N-14)}{60}$$

for runs of length k for $k < N-1$,

$$\frac{2[(k^2+3k+1)N-(k^3+3k^2-k-4)]}{(k+3)!}$$

for runs of length N - 1,

$$\frac{2}{N!}$$

(b) Hypothesis. If the residuals are truly randomly distributed, the number of runs up and down will (statistically) match the expected values.

H_o = The residual sequence is randomly distributed.

H_a = The residual sequence is not randomly distributed.

(c) Statistical Procedures. As described above, the sequence of residual values for each model were converted into binary sequences, S. Then the runs of length k were counted. These "counted" values were then compared to the expected values which were calculated according to the formulas above. A χ^2 test of goodness of fit was then applied to check the acceptability of the randomness at a 95 percent level of significance.

Runs Above and Below the Mean Test

(a) Statistical Theory. The Runs Above and Below the Mean Test theory closely mirrors the theory discussed above for the Runs Up and Down Test. First, the sequence is normalized to the range (0,1) with a mean of 0.5. Then, each residual value r_i is compared to the

mean and assigned a value of *zero* if $r_i < 0.5$ and a value of *one* if $r_i > 0.5$ (21:61). The runs in sequence S can then be counted and compared to the expected values. The expected number of runs of length k is

$$(N - k + 3)2^{-k-1}$$

and the expected total number of runs is

$$\frac{(N+1)}{2}$$

Again, a χ^2 goodness of fit test is used to check the acceptability of the randomness of the residual sequence.

(b) Hypothesis. If the residual sequence is truly random, the number of runs above and below the mean will (statistically) match the expected values.

H_o = The residual sequence is randomly distributed.

H_a = The residual sequence is not randomly distributed.

(c) Statistical Procedures. A binary sequence, S, was produced from the residual sequence as discussed in the theory section above. Then, the runs of length k in sequence S were counted. The counted values were subsequently compared to the expected values which were calculated as shown above. A χ^2 goodness of fit test was then used to check the acceptability of the randomness at a 95 percent confidence level.

Frequency Test

(a) Statistical Theory. The frequency test checks the uniformity of a sequence of M consecutive sets of N pseudorandom numbers. The pseudorandom numbers $r_1, r_2, ..., r_N$ are divided into x equal subintervals within the (0,1) unit interval. Accordingly, the *expected*

number of random numbers in each subinterval is N/x. The *actual* number of pseudo-random numbers r_i $(i = 1, 2, ..., N)$ in the subinterval $(j-1)/x \leq r_i < j/x$ is defined as f_j, where $j = 1, 2, ..., x$ (21:58). If the sequence is "truly" random, the statistic χ_1^2 has approximately a chi-square distribution with $x - 1$ degrees of freedom. The χ_1^2 statistic is then computed for all M consecutive sets of N pseudorandom numbers according to:

$$\chi_1^2 = (\frac{x}{N}) \sum_{j=1}^{x} (f_i - \frac{N}{x})^2$$

The number of resulting M values of χ_1^2 which lie between the $(j-1)$th and jth quantile of a χ^2 distribution with $x-1$ degrees of freedom $(j = 1, 2, ..., u)$ is defined as F_j (21:58). A new test statistic χ_F^2 can be computed from

$$\chi_F^2 = \frac{u}{M} \sum_{j=1}^{u} (F_j - \frac{M}{u})^2$$

.

The χ_F^2 test statistic can then be compared to a $\chi_{(.95)}^2$ critical value.

(b) Hypothesis. If the pseudorandom sequence is truly random, the χ_F^2 test statistic will not exceed the $\chi_{(.95)}^2$ critical value.

(c) Statistical Procedures. The residual sequence for each model was divided into intervals ($x = 9$) and the χ_F^2 test statistic was calculated according to the formula listed above. The χ_F^2 test statistic was then compared to the $\chi_{(.95)}^2$ critical value to see if the randomness criterion was met.

Portmanteau Test

(a) Statistical Theory. The Portmanteau test uses the first 20 autocorrelations of the residuals, *taken as a whole*, as an indicator of inadequacy of a model. If the fitted model is

appropriate, then the Portmanteau test statistic, Q, (which is equal to the summation of the first K autocorrelations, $r_k(\hat{a})$ $(k = 1, 2, ..., K)$ from any ARIMA model) will be approximately distributed as a $\chi^2(K-p-q)$, where $n = N-d$ is the number of weights, w, used to fit the model (3:291).

$$Q = n \sum_{k=1}^{K} r_k^2(\hat{a})$$

The adequacy of the model may be assessed by comparing the Portmanteau test statistic, Q, to the $\chi^2_{.95}$ value.

(b) Hypothesis. If the model is adequate, the Q values will be less than the $\chi^2_{.95}$ values for the first 20 lags. Conversely, if the model is inadequate, at least one of the first 20 lag Q values will exceed the $\chi^2_{.95}$ value.

(c) Statistical Procedures. The residual values for the first 20 lags were used to compute the Portmanteau test statistics as discussed above. These values were subsequently compared to the $\chi^2_{.95}$ values to assess the adequacy of the model. The model was considered to "pass" if all 20 Q values were less than the $\chi^2_{.95}$ values, and "failed" if any Q value exceeded the $\chi^2_{.95}$ value.

(AIC) Test

(a) Statistical Theory. In choosing a model, the more terms used in the model (higher values of p and q) the better fitted the model will be to the observed data. However, "over-fitting" the model (by tailoring the model parameters p and q to the data) can result in grossly erroneous predictions (5:287). Akaike recognized this dilemma and developed a criterion which assigns a cost for the introduction of each additional parameter above

the minimum necessary to reduce the residuals to "white" noise. Akaike's Information Criterion (AIC) is used to compare feasible models[10] (5:287).

$$AIC = k \ln \sqrt{\sigma} + 2(\frac{k}{n})(p + q + S_p + s_q + 2 - \text{Sign}(d + S_d)) + k \ln 2\pi + k$$

where,

σ = the standard deviation,

k = the number of lags,

n = the total number of observations,

p = the autoregressive term,

q = the moving average term,

S_p = the "seasonal" autoregressive term,

S_q = the "seasonal" moving average term,

d = the number of differences.

(b) Hypothesis. The AIC test is used as a means of comparing feasible models. The lowest AIC value represents the most parsimonious (and therefore "best") model.

(c) Statistical Procedures. The AIC value for each feasible model was calculated according to the formula shown above. The AIC values were then compared to determine the "best" model for each cardinal direction at Columbia and Kirtland.

The first four tests are used to eliminate infeasible models.[11] The remaining (feasible) models are then subjected to the Akaike Information Criterion (AIC) test. Akaike developed a criterion which assigns a cost for the introduction of each additional parameter (term) in a model (5:287). Thus, the best model is the model that adequately estimates the data, yet uses

[10] Feasible models are defined here to be models which have been tested to ensure the residuals are random. In our application, we use the Runs tests, Portmanteau test, and Frequency test to meet this criterion.

[11] Infeasible refers to models that do not adequately estimate the original data set.

the least number of terms. The best model is indicated by the lowest value of the AIC (5:287). The results of the five tests of residuals for the "best" model at each site are presented in Appendix G.

Full Resolution versus Grid Subset Tests

- Statement of Research Question. Lund and Shanklin used a subset of data from their photos to conduct their analysis, but never statistically demonstrated that the subset of data accurately represents the full resolution image. The second research objective is to statistically evaluate the correlation between the WSI full-resolution images (full images taken every 10 minutes) and the Lund and Shanklin "template"-generated subset.

  Full image data is archived on WSI tapes once every ten minutes.[12] Additionally, a subsample of the full image is taken at one-minute intervals. This subsample consists of the pixels that form a 33 × 33 grid on the full image (see Figure 3.6). These images are referred to as the *grid* subset.

  Our approach to verify that the Lund and Shanklin (L&S) subset accurately represents the full image will be conducted in three phases.

  1. Phase I: Correlate the L&S subset to the grid subset.
  2. Phase II: Correlate the grid subset to the full image. Some preliminary work has already been performed to establish the correlation between the grid subset and the full image (Phase II). When the WSI system was under development, Janet Schields from Scripps Institution of Technology conducted correlation tests to determine if the proposed grid format would be representative of the full image. She compared the percentage of cloud cover from the grid subset to the cloud cover in the full image. From Shields' study, the correlation between the grid and the full image is above 98 percent (26). However,

[12]See Appendix A for more information on the data collection process.

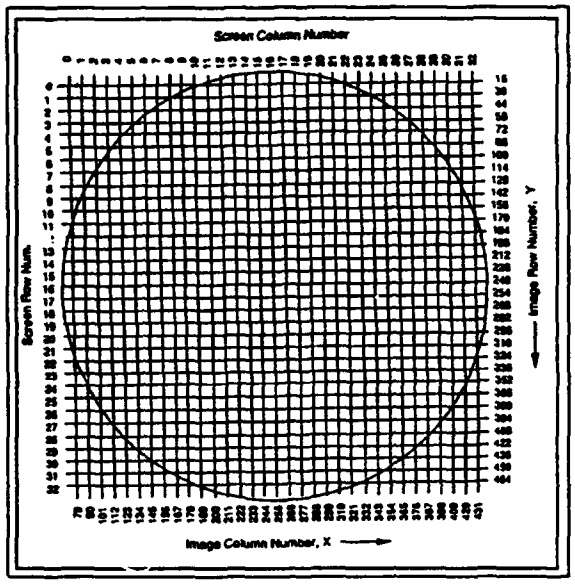

Figure 3.6. Format of the grid superimposed over the full image.

these results were based on a relatively small sample size. Therefore, a more robust examination of a statistically-significant sample of images will be required to firmly establish the correlation between the grid subset and the full image.

3. Phase III: Develop a relationship between the L&S subset and the full image by coupling the two established correlations of Phases I and II.

- Statistical Model and Theory. The issue of whether or not the Lund and Shanklin subset adequately represents the full image is closely related to the issue of azimuthal independence. The important point to remember is how Lund and Shanklin used the data and what they inferred from the data.[13]

  As a reminder, Lund and Shanklin were developing a model which determined the PCFLOS as a function of elevation angle and percentage of total sky cover. The percentage of total sky cover was based on readings from NWS observers. The other parameter that Lund and Shanklin needed was whether or not it was clear or cloudy at given elevation angles.

[13] Another point to bear in mind is that Lund and Shanklin did not use the data to infer any spatial distribution within the sky-dome. Accordingly, spatial analyses would be inappropriate tests.

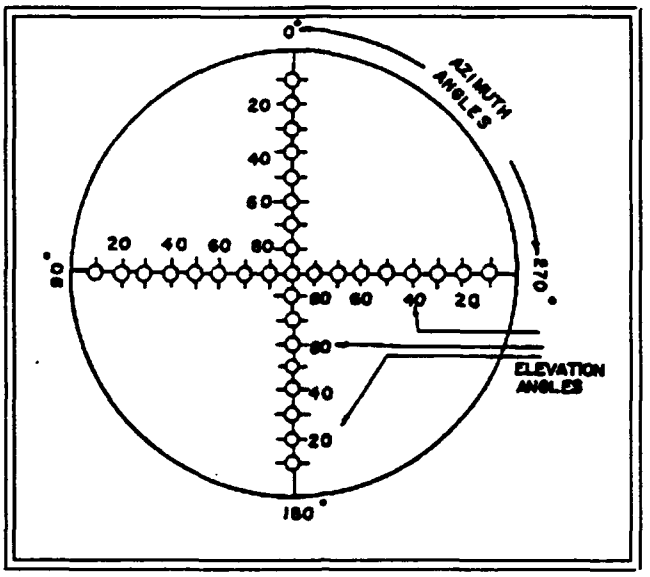

Figure 3.7. The Lund and Shanklin template.

To empirically develop the needed statistics, Lund and Shanklin developed a template to subsample the sky-dome (see Figure 3.7).

Each elevation angle is represented by an annular ring. In essence, Lund and Shanklin's template is a subset that uses only four points off of the annular ring for each elevation angle. On face value, this is a rather severe subsampling. For example, the 30° annular ring consists of over 300 pixels. But, if the assumption of azimuthal independence is absolutely true, then knowing just one value on each annular ring should be sufficient to represent the values of the entire ring. However, we know from our earlier investigation of Research Question One that azimuthal independence is not an absolute property. Therefore, we will test to see if the four points of Lund and Shanklin's template that lie on each annular ring adequately represent that ring. Likewise, we will test to see if the pixels from the grid subset that lie on each annular ring represent that ring. The time series analysis methodology developed for azimuthal independence testing will be employed to test if the subsets adequately represent the larger images.

- **Hypothesis.** By applying the time-series analysis developed on page 3-8, we can deduce the best ARIMA model to represent each elevation angle. Our hypothesis is that the "best" model as determined by using all the data points on the annular ring at a given elevation angle will be the same "best" model as determined by only using the four points from the Lund and Shanklin template at that elevation angle. If this hypothesis is true, then the four points used by Lund and Shanklin are adequate to represent the observations for the given elevation angle. Conversely, if a different model is indicated, then the inclusion of all the data points on the annular ring must be the cause of the differing models. This, of course, implies that the four data points used in the Lund and Shanklin template did not adequately represent the observations at the given elevation angle.

- Model Assumptions, Limitations, and Applicability. The same assumptions and limitations discussed for time-series analysis on 3-10 apply here as well. Due to the interrelated issue of azimuthal independence, the time-series analysis methodology developed to test azimuthal independence is also an appropriate methodology to use here.

- Statistical Procedures. The methodology and statistical procedures for evaluating Phases I and II are the same, so only Phase I will be described in detail. The appropriate substitutions and analogies to conduct Phase II are straightforward.

  To evaluate Phase I, all the pixels that comprise the annular rings at each elevation angle (10° to 70° in 10° increments) are extracted and evaluated to determine an observation of clear, thin-cloud, or thick-cloud. Then, the observations of CFLOS (at a constant elevation) for an entire month are added. The percentage of CFLOS over the course of 15 months can then be plotted by concatenating the monthly percentages together. This "pseudo" time series is analogous to the time-series developed on page 3-10. Once the "pseudo" time series is developed, a complete time-series analysis can be conducted by following the procedures whic begin on page 3-10. The results will be a "best" ARIMA model for each elevation angle.

These models can be compared to the results obtained by only using the four points on each annular ring that are intercepted by Lund and Shanklin's template.[14]

PCFLOS Model Estimates versus WSI Database Observations Tests

- Statement of Research Question. The third research objective is to determine the correlation between the probability estimates as determined by the Lund and Shanklin and SRI models and the actual WSI database observations.
- Statistical Model and Theory. We chose to use the SRI model for comparison with the WSI observations since the SRI model is a refinement of the Lund and Shanklin model and is generally accepted as the more accurate of the two theoretical models.

  The SRI model will be used to generate PCFLOS theoretical estimates by varying the elevation and total sky cover parameters as follows:

  - Elevation. The elevation will be varied in 10-degree increments from 0 to 70 degrees from zenith.
  - Total Sky Cover. The total sky cover variable will range from 0 percent sky cover to 100 percent sky cover in 10-percent increments.

  These estimates will be calculated and plotted for comparison against the WSI observations from the Columbia and Kirtland sites.
- Model Assumptions, Limitations, and Applicability. No statistical analysis is conducted in comparing the theoretical estimates and actual WSI observations.
- Statistical Procedures. Each WSI one-minute image will be sorted according to the percent total sky cover (in tenths). Then, PCFLOS estimates will be calculated by summing and taking the percentage of clear observations for each elevation angle. This data will then be

[14] These results were already completed in the azimuthal independence testing.

plotted and overlaid on the theoretical estimates developed above. These combined plots will then be subjected to visual inspection. Additionally, the differences between the theoretical estimates and the WSI observations will then be tabulated by elevation angle and percentage of total sky cover to portray the differences more objectively.

IV. Analysis and Results

The purpose of this chapter is to present the results of the tests outlined in Chapter III. This chapter is organized into three subsections—one for each research question.

Azimuthal Independence Tests

χ^2 Test of Proportionality. **Our rejection criterion was to reject H_o if χ^2 >12.59, the value of $\chi^2_{0.05}$ for (3-1)(4-1)= 6 degrees of freedom.**

The results of the individual tests for proportionality have been tabulated and are displayed in Table 4.1.

Table 4.1. Results of χ^2 statistics (by month and elevation) for Columbia Missouri, February 1989 to January 1990.

Month	*Elevation*			
	10°	30°	50°	70°
Feb 89	7.9	37.6	198.4	219.8
Mar 89	27.9	96.7	175.3	173.2
Apr 89	51.3	145.2	304.6	347.2
May 89	37.4	93.2	98.5	107.9
Jun 89	46.0	128.0	28.1	26.0
Jul 89	182.4	267.1	283.1	227.3
Aug 89	28.7	263.3	500.3	1588.4
Sep 89	12.9	159.3	304.4	386.5
Oct 89	76.6	368.1	526.9	752.9
Nov 89	58.7	204.1	473.7	432.2
Dec 89	43.7	82.1	142.7	138.1
Jan 90	539.6	1727.9	2395.0	1068.23
CUM	573.2	2264.5	4162.8	3868.7

Of the 52 individual tests conducted, only the February 1989 at 10 degrees is in the acceptance region at the 0.05 level of significance. These results lead us to reject the null hypothesis. This outcome is not surprising though. As pointed out by Devore (1991), a large sample will almost always lead to rejection of the null hypothesis. This results because any small departure from the null hypothesis will be detected by the test, yet will be of little practical significance (7:320). For example, we tested the hypothesis that the proportion of observations are the same (thus,

independent of azimuth). In effect, we were testing to see if the percentage of observations in each direction were the same—*exactly the same.* The first contingency table in Appendix D demonstrates this point rather clearly. The percentage of clear observations over the entire year in each direction (east, north, south, west) are 53.83, 52.19, 52.80, and 53.80, respectively. Due to the large sample size, the χ^2 test appropriately discerns that $53.83 \neq 52.19 \neq 52.80 \neq 53.80$ and leads us to reject the null hypothesis. But, the differences in the percentages are trivial for our purposes. Devore (1991) refers to this dilemma as the difference between *statistical* significance and *practical* significance. What we are really interested in is the closeness of the proportion of observations that in each direction. The example we just used from Appendix D represents the cumulative results over the entire year. And, as have already noted, the percentages of clear observations were extremely close in each direction. Upon further examination, we found the same closeness of percentage of clear observations in each direction for the data accumulated at monthly intervals (regardless of the elevation angle). Table 4.2 below displays the 30° elevation angle percentages of clear observations as a representative sample. The monthly contingency tables used to create Table 4.2 are included in Appendix D for completeness.

Table 4.2. Percentage of clear observations for each direction at Columbia (by month).

Month	*Direction*					
	East	North	South	West	$\overline{X}$	σ^2
Feb 89	45.03	47.12	45.24	45.98	45.84	0.67
Mar 89	48.26	51.36	48.57	49.42	49.43	1.46
Apr 89	47.20	52.63	49.13	50.30	49.85	3.87
May 89	39.64	39.73	39.66	38.77	39.46	0.16
Jun 89	49.62	50.70	53.74	55.93	52.38	6.21
Jul 89	36.67	31.58	34.70	30.92	33.36	5.46
Aug 89	48.54	51.81	51.45	45.24	49.26	6.99
Sep 89	55.72	57.87	57.04	54.58	56.26	1.58
Oct 89	66.64	66.39	64.80	60.92	64.69	5.23
Nov 89	74.53	74.19	70.35	69.62	72.17	4.87
Dec 89	69.72	66.71	61.88	61.05	64.84	12.61
Jan 90	73.03	67.22	63.61	61.87	66.43	18.23

With only a few exceptions, the percentage of clear observations for each direction are within two standard deviations of the mean for each month.

Based purely on *statistical* significance, we would correctly deduce from the χ^2 tests that the null hypothesis should be rejected. However, it would be erroneous to conclude that this meant the proportion of observations are not independent of azimuth. It simply means that the proportion of observations are not *statistically* the same. From a perspective of *practical* significance, the χ^2 tests demonstrate that the proportion of clear[1] observations in each direction are virtually the same. Since we are primarily interested in whether or not the proportions are approximately the same it would be correct to use the *practical* significance perspective.

Trend Analysis. The purpose of this test was to provide an indication as to the correlation between 1) the observations at different elevation angles along the same axis; and, 2) the correlation between each axis.

Correlation at Different Elevation Angles. As depicted in Figure 4.1, a strong correlation exists between observations at elevation angles from 10° to 60° along the same cardinal axis.[2] (This observed correlation will allow us to use *spatial* versus the usual *temporal* differencing techniques to develop a stationary data set in our ARIMA model testing below.)

Correlation Between the Cardinal Directions. The percentage of observations that are the same along each cardinal axis are plotted versus the lags in Figure 4.2. To support the null hypothesis, each of these profiles should show a close correlation.

As can be noted in Figure 4.2, this is not the case. Specifically, the east profile is lower than the north and west; the south profile has an exaggerated standard deviation (by comparison); and, the slopes of the profiles in each direction are different. All of these differences indicate that the

[1] By extension, the same holds true for thin-cloud and thick-cloud observations.
[2] The strong correlation degenerates past 60°, so the 70° and 80° trend lines were omitted from these plots.

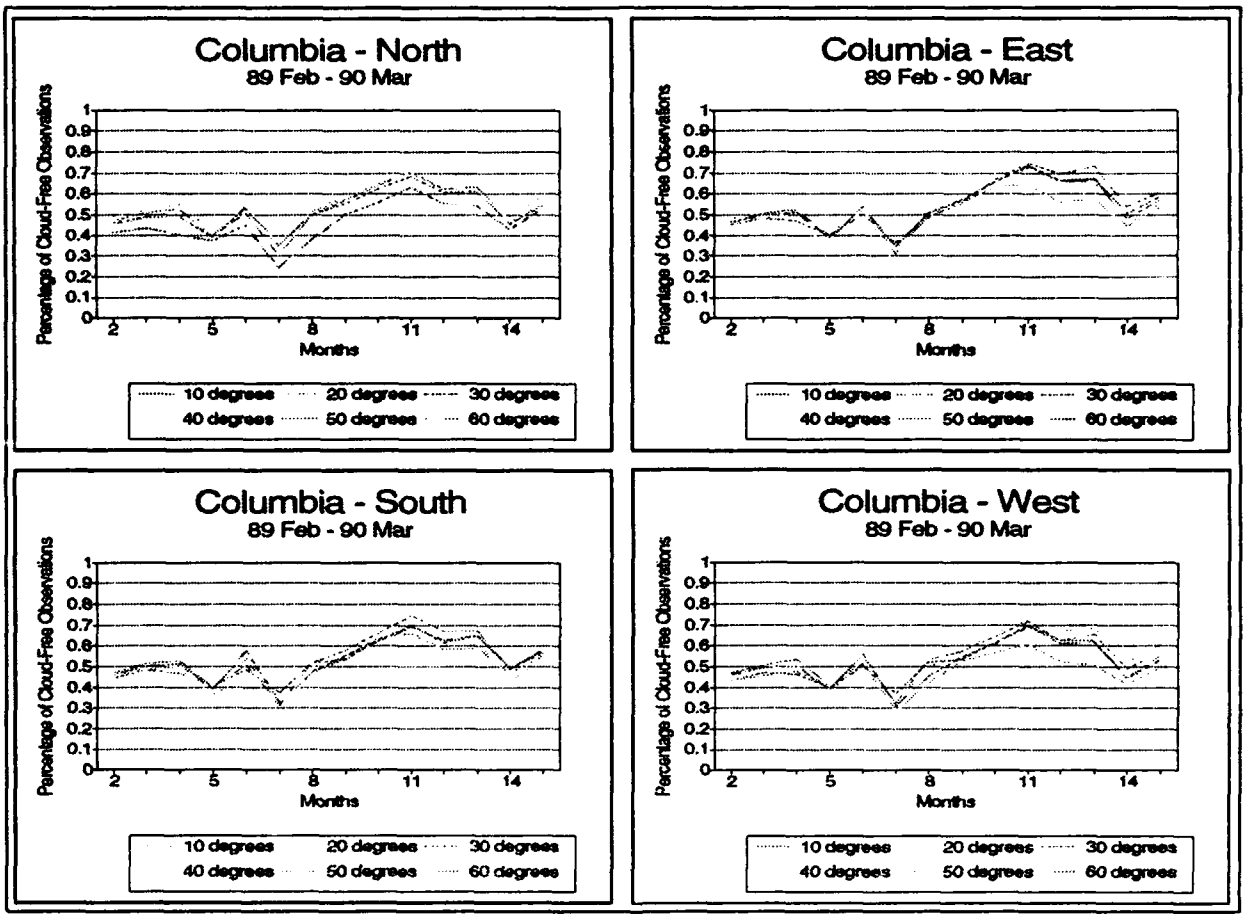

Figure 4.1. Plots of the percentage of clear days for Columbia, Missouri from February 1989 to January 1990.

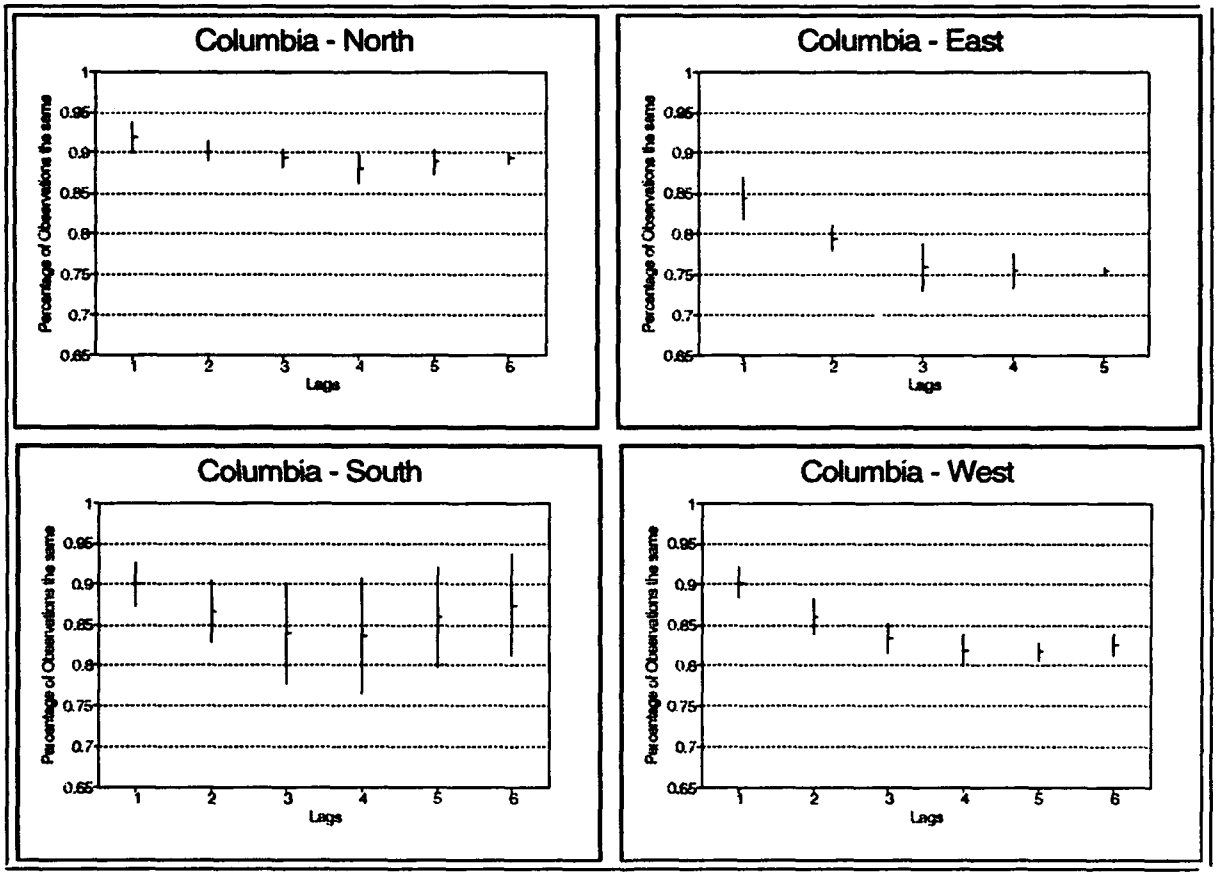

Figure 4.2. Plots of the one standard deviation profiles for percentage of same readings for each direction.

rate of change of the observations is not the same along each cardinal direction. Accordingly, we subjectively reject the null hypothesis based on the differences in the profiles in Figure 4.2. However, these apparent differences may be attributed to other phenomenon such as sun movement and solar occulter interference. The fact that the east and west plots in Figure 4.2 are fairly close may be attributable to the symmetry of the sun's motion (and solar occulter motion) in the east-to-west direction. Also, the larger variance of the south's profile may be a function of the larger percentage of missing data and higher percentage of false readings due to the continuous presence of the occulter and the Sun's effect on the nearby sky. Due to these uncertainties and the lack of a rigorous statistical test, little significance is placed on the results of this test.

Time Series Analysis. The plots of the ACFs, PACFs, and periodograms were used to identify candidate models. As can be seen by examining the plots in Figures F.9 through F.16, the PACFs have a characteristic dominant first lag term and the ACFs all follow a damped exponential pattern. These both indicate an ARMA (100) model. Also, the periodograms show a high density in the low end of the frequency spectrum which indicates the periodicity has not yet been removed. So, when we estimate the models, we should expect the ARMA (100) models to closely represent the data and have residuals that have been reduced to "white" noise. Some of the PACF plots (for example, Kirtland-North) show significant spikes at lags less than five that may also prove to be suitable models.

Normally, only the candidate models would be estimated and subjected to further testing. But, we opted to estimate all the models since we had the computer capacity required.[3] After the models were estimated, the ACF, PACF, cumulative probability plots, and periodograms of the residuals of each model were plotted. These plots are included in Appendix F for completeness.

[3]However, from our identification phase results, we would expect that the ARMA (100) models would prove to be good, if not the best, models.

We estimated each model for both sites and all four directions using Statgraphics®subroutines. When a model is estimated, the autoregressive (p) terms and moving average (q) terms reduce the original series to a residual series. In an ARIMA model that adequately represents the data, the residuals are reduced to "white" noise. The residuals were subjected to visual inspection and quantitative tests to ensure they were reduced to white noise.

1. Visual Inspection of Plots. The ACFs, PACFs, normal probability plots, and periodograms displayed in Appendix F were visually examined. For a well-fitted model, the ACF and PACF plots would be reduced to small[4] random spikes which are indicative of white noise. Also, for white noise, the normal probability plot would be scattered points about a line joining (0,0) and (0.5,1) (3:295). The periodogram for a well-fitted model would approximate a uniform distribution. Models that represented the data well would produce well-behaved plots and models that used the wrong AR or MA terms would show anomalous spikes and vary from the desired "white" noise profiles described above. From our examination of the ACFs and PACFs of the original data, we expect the ARIMA (1,0,0) plots to be the best behaved.

   Upon examination, it became apparent that the plots for each model possessed the same characteristic patterns regardless of direction. Additionally, the AR and MA terms dominated entire groups of models. This enabled us to classify the models into three distinct categories. The categories and their trends are enumerated below:

   (a) Models without an AR term: (0,0,1), (0,0,2).

   - ACF and PACF Plots. The ACF and PACF plots routinely violated the standard error criteria in lags one and two for the (0,0,1) model and at lag three for the (0,0,2) models.
   - Normal Probability Plots and Periodograms. The normal probability plots appeared to diverge at the tails. The periodograms characteristically exhibited a spike in the

[4] "Small" here refers to the spike being confined within plus or minus twice the standard error.

low frequency end of the spectrum which indicates that all the significant AR and MA terms were not accounted for in the model.

(b) Models with an AR term and no MA term: (1,0,0), (2,0,0).

- ACF and PACF Plots. With the exception of an outlier at Columbia-East, these plots appeared well-behaved and generally contained within the standard error limits. The (2,0,0) models very closely mirrored the (1,0,0) models, but had slightly larger spikes.
- Normal Probability Plots and Periodograms. The normal probability plots were extremely difficult to differentiate between and thus provided little information in comparing the models. The periodograms for both the (1,0,0) and the (2,0,0) models consisted of uniformly distributed spikes which indicates that the models are adequate.

(c) Models with mixed AR and MA terms (1,0,1), (1,0,2), (2,0,1), (2,0,2).

- ACF and PACF Plots. The (1,0,1) and (1,0,2) models showed very small spikes at low lag values, but tended to increase and ultimately violate the standard error limits at higher lags. Additionally, the PACF plots appeared less random and appeared to follow a weak exponential pattern. The (2,0,1) and (2,0,2) models exhibited the same small residual spikes as the (2,0,0) model. Furthermore, the addition of the MA terms appeared to reduce the size of the spikes. As a result, the (2,0,2) model exhibited the best behaved combination of ACF and PACF plots.
- Normal Probability Plots and Periodograms. Again, the normal probability plots were extremely difficult to differentiate between and were not used in comparing the models. The periodograms for this category of models exhibited generally uniform distributions which indicated the models fit the data well.

Based on the visual examination of the plots, the (2,0,2) model appears to produce the best behaved residuals. The residuals of the (1,0,0), (2,0,0), and (2,0,1) models also appear to be reasonably small and well-behaved. The (0,0,1) and (0,0,2) models did not produce random residuals. The poor fit of these models further indicates the dependence on the presence of an AR term to adequately represent the data.

2. Quantitative Tests of Residuals. After the visual inspection of the residuals, each model was subjected to five quantitative tests:

   (a) Runs Up and Down Test (**RUD**)

   (b) Runs Above and Below the Mean Test (**RABM**)

   (c) Frequency Test (**FREQ**)

   (d) Portmanteau Test (**PORT**)

   (e) Akaike Information Criterion (**AIC**)

The results of the tests of residuals on each model are summarized in Tables 4.3 through 4.10 below. Additionally, an expanded version of the results for the "best" model at each site and direction are displayed in Appendix G.

Table 4.3. Summary of the results of the tests of residuals of Columbia-North data.

Model	RUD	RABM	FREQ	PORT	AIC
CN(001)	F	P	P	F	N/A
CN(002)	P	P	P	P	-449.79
CN(100)	P	P	P	P	-473.29
CN(101)	P	P	P	P	-471.30
CN(102)	P	P	P	P	-471.50
CN(200)	P	P	P	P	-471.30
CN(201)	P	P	P	P	-469.64
CN(202)	P	P	P	P	-469.65

As expected from the identification phase, the ARMA (100) model proved to be the "best" model for six of the eight possible directions. The only exceptions were the Columbia-East and

Table 4.4. Summary of the results of the tests of residuals of Columbia-East data.

Model	RUD	RABM	FREQ	PORT	AIC
CE(001)	F	P	P	P	N/A
CE(002)	P	P	P	P	-376.26
CE(100)	F	P	P	P	N/A
CE(101)	F	P	P	P	N/A
CE(102)	P	P	P	P	-395.99
CE(200)	F	P	P	P	N/A
CE(201)	F	P	P	P	N/A
CE(202)	P	P	P	P	-396.33

Table 4.5. Summary of the results of the tests of residuals of Columbia-South data.

Model	RUD	RABM	FREQ	PORT	AIC
CS(001)	F	P	P	P	N/A
CS(002)	P	P	P	P	-413.22
CS(100)	P	P	P	P	-423.11
CS(101)	P	P	P	P	-422.76
CS(102)	P	P	P	P	-422.15
CS(200)	P	P	P	P	-423.29
CS(201)	P	P	P	P	-421.82
CS(202)	P	P	P	P	-420.52

Table 4.6. Summary of the results of the tests of residuals of Columbia-West data.

Model	RUD	RABM	FREQ	PORT	AIC
CW(001)	P	P	P	P	-402.10
CW(002)	P	P	P	P	-410.34
CW(100)	P	P	P	P	-417.93
CW(101)	P	P	P	P	-415.97
CW(102)	P	P	P	P	-413.89
CW(200)	P	P	P	P	-415.99
CW(201)	F	P	P	P	N/A
CW(202)	P	P	P	P	-411.95

Table 4.7. Summary of the results of the tests of residuals of Kirtland-North data.

Model	RUD	RABM	FREQ	PORT	AIC
KN(001)	F	P	P	P	N/A
KN(002)	P	F	P	P	N/A
KN(100)	P	P	P	P	-414.99
KN(101)	P	P	P	P	-413.43
KN(102)	P	P	P	P	-358.19
KN(200)	P	P	P	P	-413.18
KN(201)	P	P	F	P	N/A
KN(202)	P	P	F	P	N/A

Table 4.8. Summary of the results of the tests of residuals of Kirtland-East data.

Model	RUD	RABM	FREQ	PORT	AIC
KE(001)	F	P	P	P	N/A
KE(002)	P	P	P	P	-347.66
KE(100)	P	P	P	P	-360.97
KE(101)	P	P	P	P	-359.99
KE(102)	P	P	P	P	-358.19
KE(200)	P	P	P	P	-359.87
KE(201)	P	P	P	P	-357.52
KE(202)	P	P	P	P	-357.01

Table 4.9. Summary of the results of the tests of residuals of Kirtland-South data.

Model	RUD	RABM	FREQ	PORT	AIC
KS(001)	F	P	P	P	N/A
KS(002)	P	P	P	P	-347.29
KS(100)	P	P	P	P	-360.39
KS(101)	P	P	P	P	-358.90
KS(102)	P	P	P	P	-358.14
KS(200)	P	P	P	P	-359.09
KS(201)	P	P	P	P	-356.38
KS(202)	P	P	P	P	-356.26

Table 4.10. Summary of the results of the tests of residuals of Kirtland-West data.

Model	RUD	RABM	FREQ	PORT	AIC
KW(001)	P	P	P	P	-405.42
KW(002)	P	P	P	P	-405.72
KW(100)	P	P	P	P	-407.06
KW(101)	P	P	P	P	-406.09
KW(102)	P	P	P	P	-405.30
KW(200)	P	P	P	P	-405.52
KW(201)	P	P	P	P	-404.39
KW(202)	P	P	P	P	-403.68

South data which were best modeled as an ARMA (202) and ARMA(200), respectively. Further discussion of these results is contained in the conclusions section of Chapter V.

Full Resolution versus Grid Subset Tests

This analysis was not performed due to the non-availability of the WSI 10-minute data tapes required to perform the Phase II analysis.

PCFLOS Model Estimates versus WSI Database Observations Tests

The PCFLOS curves generated by the SRI model are shown as dashed lines in Figures 4.3 and 4.4 for the Columbia and Kirtland data, respectively. PCFLOS curves based on the observations in the WSI database are superimposed as solid lines. As can be seen from a visual inspection, the SRI model estimates and WSI observations follow the same trend patterns and correlate fairly well. Tables 4.11 and 4.12 show the actual Δs between the SRI model and WSI observation estimates according to elevation angle and percent total sky cover. Note that the Δs are predominantly very small (in the hundredths) except at 80° off-zenith. Two possible explanations for the discrepancies at the 80° elevation angle are:

1. The WSI system could have trouble discriminating between cloudy and clear conditions at the horizon due to forward scattering or some other atmospheric phenomenon.

2. The theoretical model may not account for some parameter that effects the PCFLOS estimate more severely near the horizon.

A second observation is that the Δs are smaller for the Columbia data. This is not too surprising though, considering the model was developed from photographs taken at the Columbia site.

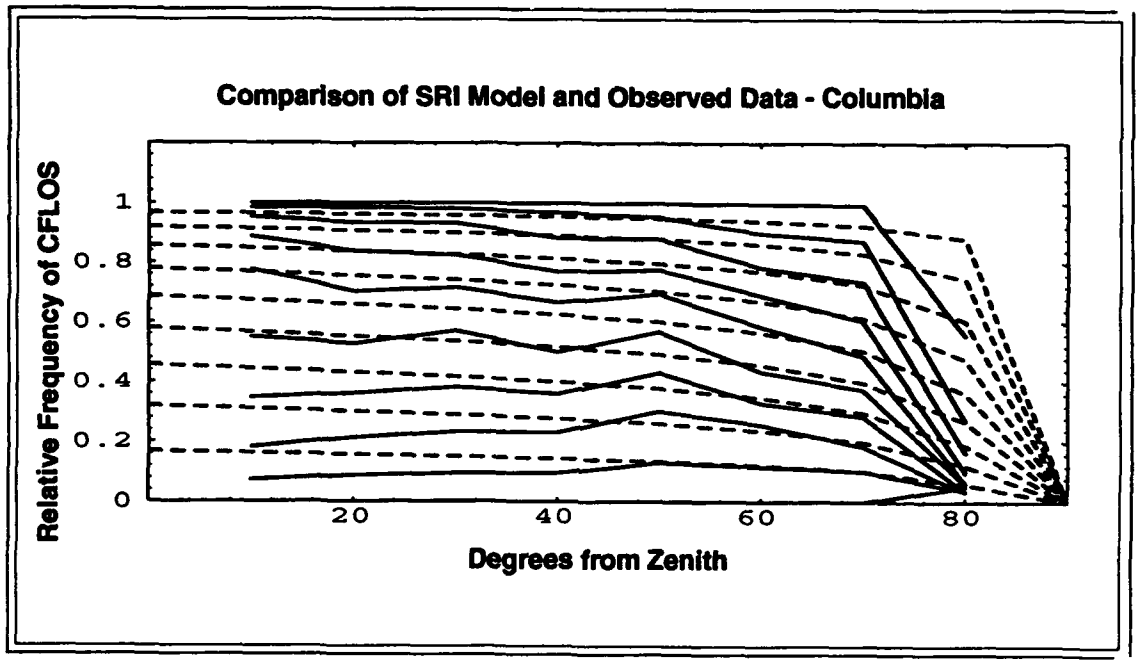

Figure 4.3. PCFLOS estimates of the SRI model (dashed) versus the WSI observed data for the Columbia site. (The top line represents the ten percent total sky cover condition, the second line from the top represents the twenty percent total sky cover condition, ...)

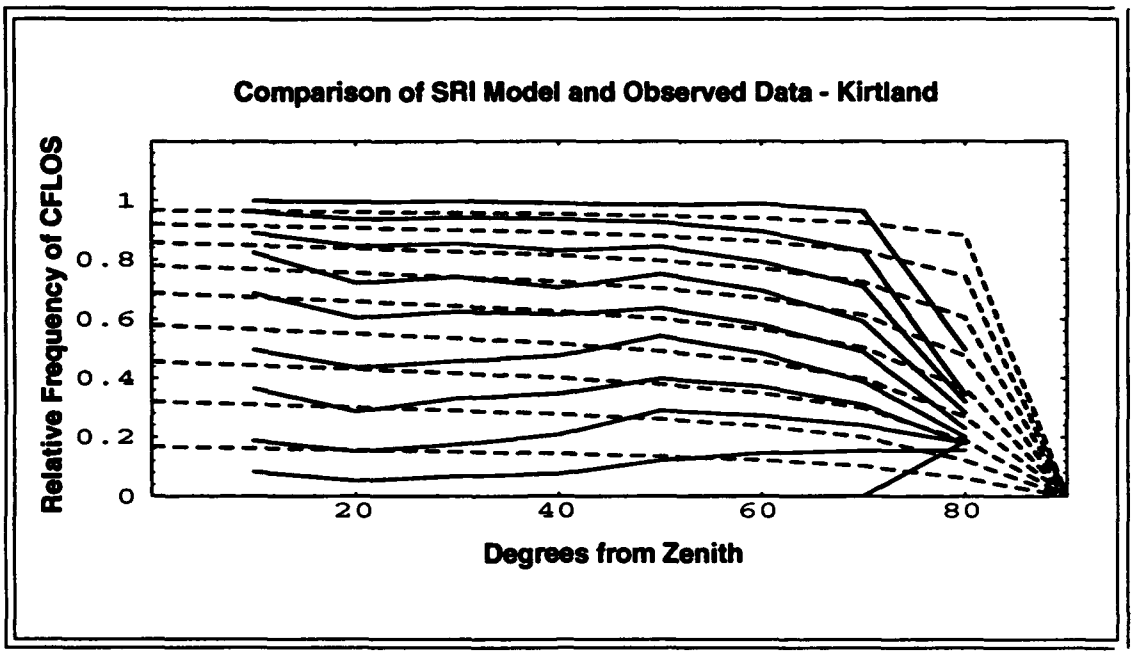

Figure 4.4. PCFLOS estimates of the SRI model (dashed) versus the WSI observed data for the Kirtland site. (The top line represents the ten percent total sky cover condition, the second line from the top represents the twenty percent total sky cover condition, ...)

Table 4.11. Δs between the SRI model and WSI observation estimates in the probability of cloud-free lines-of-sight at the Columbia site as a function of elevation angle and observed total sky cover.

		Degrees Off-Zenith							
		10	20	30	40	50	60	70	80
	10	.00022	.00041	.00055	.00037	.00056	.00073	.00052	.05221
	9	.08950	.06922	.05582	.04786	.00386	.00505	.00005	.02188
	8	.12858	.08734	.05669	.04536	.04162	.01944	.01542	.08653
Total	7	.09741	.07042	.03519	.04070	.05218	.01865	.01329	.15033
Sky Cover	6	.01563	.02507	.03441	.01703	.07523	.02255	.02400	.20805
(tenths)	5	.10253	.04256	.07168	.04105	.09408	.01960	.01799	.29978
	4	.11923	.08283	.08348	.04473	.07254	.01960	.01001	.37828
	3	.10486	.09566	.10549	.06862	.08352	.01364	.01061	.43996
	2	.07088	.07734	.08113	.07737	.07328	.03782	.04255	.47297
	1	.03531	.03830	.04160	.04568	.05112	.05811	.06898	.32441

Table 4.12. Δs between the SRI model and WSI observation estimates in the probability of cloud-free lines-of-sight at the Kirtland site as a function of elevation angle and observed total sky cover.

		Degrees Off-Zenith							
		10	20	30	40	50	60	70	80
	10	.00025	.00016	.00025	.00044	.00043	.00157	.00110	.18490
	9	.07937	.10433	.08316	.06850	.01432	.02359	.05145	.09574
	8	.12321	.14828	.11513	.06982	.02869	.03440	.03981	.05842
Total	7	.07999	.14563	.08617	.05448	.01898	.02250	.01092	.01028
Sky Cover	6	.06863	.11411	.07683	.04089	.04999	.02725	.01113	.06737
(tenths)	5	.01187	.05469	.01872	.01055	.03598	.01520	.01427	.13375
	4	.05567	.03406	.00018	.02016	.04819	.02398	.02198	.19444
	3	.04325	.00749	.02683	.01576	.04748	.02143	.01500	.28662
	2	.04904	.02902	.04247	.04513	.04666	.03341	.00239	.39941
	1	.03464	.03335	.03950	.03647	.03708	.04929	.03940	.38279

V. Recommendations and Conclusions

Recommendations

1. We were forced to use the spatial correlation between different elevation angles instead of the usual temporal correlations when differencing the data. This problem could be eliminated by obtaining several years of data so that true *temporal* differencing can be performed. Therefore, it is recommended that several years of data be obtained and the time series analysis portion be recomputed using temporal differencing.
2. Our analysis was constrained to two WSI sites (Columbia and Kirtland) due to the non-availability of processed data from the remaining three sites. The analysis of azimuth independence should be extended to the remaining sites once the data becomes available. The incorporation of more sites from diverse geographical regions will add robustness to our conclusions.
3. Since the 10-minute data was not available, we were not able to perform the analysis to determine if Lund and Shanklin's 33-point template and the WSI system grid mesh adequately represent the full image. Once the 10-minute data set is made available, this analysis will be possible and should be performed in accordance with the methodology described in Chapter III.

Conclusions

Azimuthal Independence Tests. We applied three tests to resolve the question of azimuthal dependency.

- χ^2 Tests of Proportionality. Based on the results of the χ^2 testing, the null hypothesis (that the proportions are the same) should be rejected. However, as discussed previously, the large test statistic values are a function of the large sample size. When viewed from a perspective

of *practical* significance, the differences in the proportions are insignificant. Therefore, we conclude from the χ^2 testing that the assumption of azimuthal independence is valid.

- Trend Analysis. The trend analysis testing showed a strong correlation between observations at elevation angles from 10° to 60° along the same cardinal axis. This correlation allowed us to develop a "pseudo" time series for follow-on time series analysis testing. The trend analysis also indicated differences in the profiles of the percentage of observations that are the same along each axis. But, this was not a rigorous test and no statistical tests were employed.

- Time-Series Analysis. A strict interpretation of the results of the time series analysis, on face value, indicates that the sky-dome is independent of azimuth for the Kirtland site and is *not* independent of azimuth for the Columbia site. At the Columbia site, the ARIMA (1,0,0) model was the best model for the north and west directions. The south direction was best represented by the ARIMA (2,0,0) model and the east direction was best modeled by the ARIMA (2,0,2) model. But, after closer scrutiny, the differing models for the south and east directions can be disputed. For the Columbia-East direction, the ARIMA (1,0,0) model failed the Runs Up and Down test and was rejected as a feasible model at the 95 percent confidence level. But, by definition, we should expect to reject a true hypothesis 5 percent of the time when testing at the 95 percent confidence level (6:79). Since we tested a total of 64 combinations of sites and directions, it is not unreasonable to expect 2 or 3 of these errors. Accordingly, we are willing to accept the Columbia-East ARIMA (1,0,0) model as a feasible model[1].

Ultimately, the results of the AIC test were used to discern the "best" model in each direction. However, for the Columbia-South and East directions, the AIC values for the "best" model and other candidate models were separated by mere fractions of a point. This was

[1]This decision is also supported by the fact that the ARIMA (1,0,0) model passed all the other tests of residuals. Additionally, the original data ACF and PACF plots indicated a ARIMA (1,0,0) model for both the east and south directions and a visual inspection of the ACF and PACF plots of the residuals confirmed the similarity to the plots obtained in the other directions.

characteristically different than the results of the AIC test for the Kirtland site and for the Columbia-North and West directions. In both of these cases, the AIC value for the "best" model was decidedly lower than the next closest model (on the order of whole point values). Due to the closeness of the AIC values and the inherent sensitivity of the AIC test to small fluctuations in variance,[2] we reason that the difference in the AIC values in the Columbia south and east directions is insignificant and that either the ARIMA (2,0,0) and ARIMA (2,0,2) models *or* the ARIMA (1,0,0) model can be considered the best model.

Based on the time-series analysis, we conclude that both the Columbia and Kirtland sites are best modeled by the ARIMA (1,0,0) model in each of the cardinal directions tested. And, since the same model represents each direction, it follows that the sky-dome is isentropic.

Based on the results of the three tests of azimuthal independence discussed above we conclude that we can not reject Lund and Shanklin's assumption of azimuthal independence for the Columbia and Kirtland sites.

Full Resolution versus Grid Subset. As mentioned above, this research question has been deferred and is recommended for follow-on research efforts.

PCFLOS Estimates versus WSI Database Observations. The plots in Figures 4.3 and 4.4 and the Δs tabulated in Tables 4.11 and 4.12 both support the conclusion that the SRI model does, in fact, correlate to the WSI observations at both sites. The extent of the correlation and statistical significance has yet to be examined and is deferred for follow-on research.

Further conclusions and generalizations pertaining to the accuracy, applicability, and overall validity of the SRI model can not be made until the sub-sampling issue (Research Question II) has been resolved by follow-on research.

[2] For a given ARIMA model, the AIC value changes proportionally to the natural logarithm of the square root of the standard deviation ($AIC \propto \sqrt{\sigma}$).

Appendix A. *WSI System Data Processing(13)*

1. Hardware Description (13:8)

The purpose of this appendix is to describe the hardware used in the WSI system and explain the basic image processing from data capture through data archiving. The WSI system consists of two main parts: an exterior sensor and an interior controller unit (see Figure A.1). A further breakdown of the system components is depicted in Figure A.2.

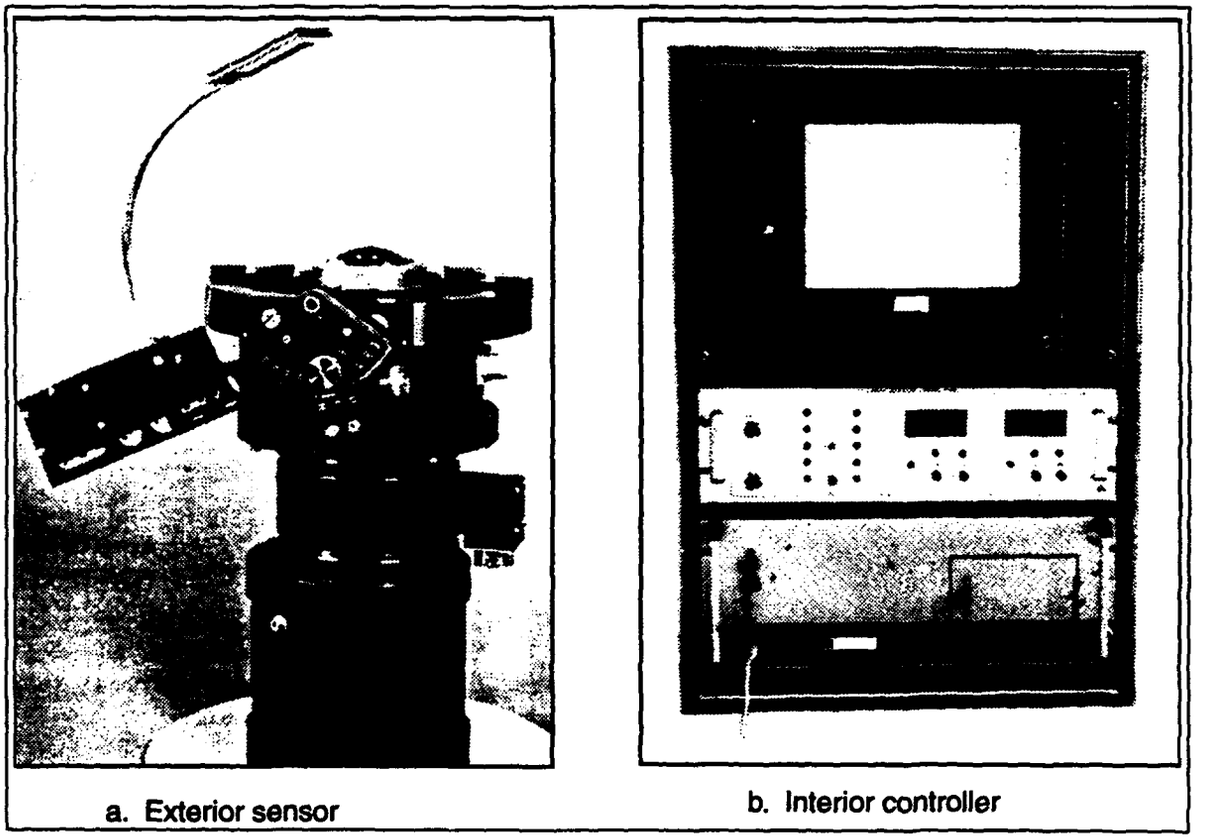

Figure A.1. The Whole Sky Imager system.

Figure A.2. Image acquisition and analysis system hardware block diagram.

The sensor is a CIDTEC Model 2710 solid-state monochromatic camera which uses a charge injection device (CID) with a 512×512 pixel array. The solar occultor, iris, and optical filters all provide control over stray light and flux variations and can be operated in an automatic

or manual mode. The optical filter mechanism contains an additional four filters for spectral band pass selection.

The interior controller is primarily comprised of a video frame grabber (1024 × 1024 image memory), the 80286-based CPU card, and an 8-millimeter cartridge. The tape cartridge provides a 2.2 gigabyte storage capability which results in a seven-day duty cycle for the entire system.

2. Data Acquisition(13:11)

The WSI system takes two seconds to capture an image. Through the use of the optical filters, the WSI system captures four images during the first eight seconds of each minute—two blue (450 nanometers) and two red (650 nanometers). These four images are digitized by an 8-bit analog to digital converter and are stored according to one of two formats. At ten-minute intervals, the entire FOV is captured and stored in a full resolution 512 × 512 matrix. At the intervening one-minute intervals, only the subset of pixels that lie on a 33 row × 33 column grid are stored. The tape formats for the one and ten-minute tapes are described in detail in a technical report prepared by Shields and are reproduced as Appendix C (24).

3. Cloud Discrimination(13:14)

After the data is stored on 8-millimeter tapes, the tapes are submitted to the Marine Physical Laboratory for calibration, format checks, and cloud discrimination processing. The basic processes involved are depicted in Figure A.3.

At this point, each of the four (two blue and two red) images has been recorded and stored in the appropriate format. To determine the presence or absence of clouds, a ratio technique is used. The blue reading is compared to the red reading (blue/red). The blue/red spectral radiance ratios are near 1.0 for white clouds and is characteristically large for clear sky conditions. After a cloud/no-cloud determination is made, an ordinal numeric value is assigned to replace each pixel reading. The values are arbitrarily set to 0, 100, 150, 200 and 250 which correspond, respectively,

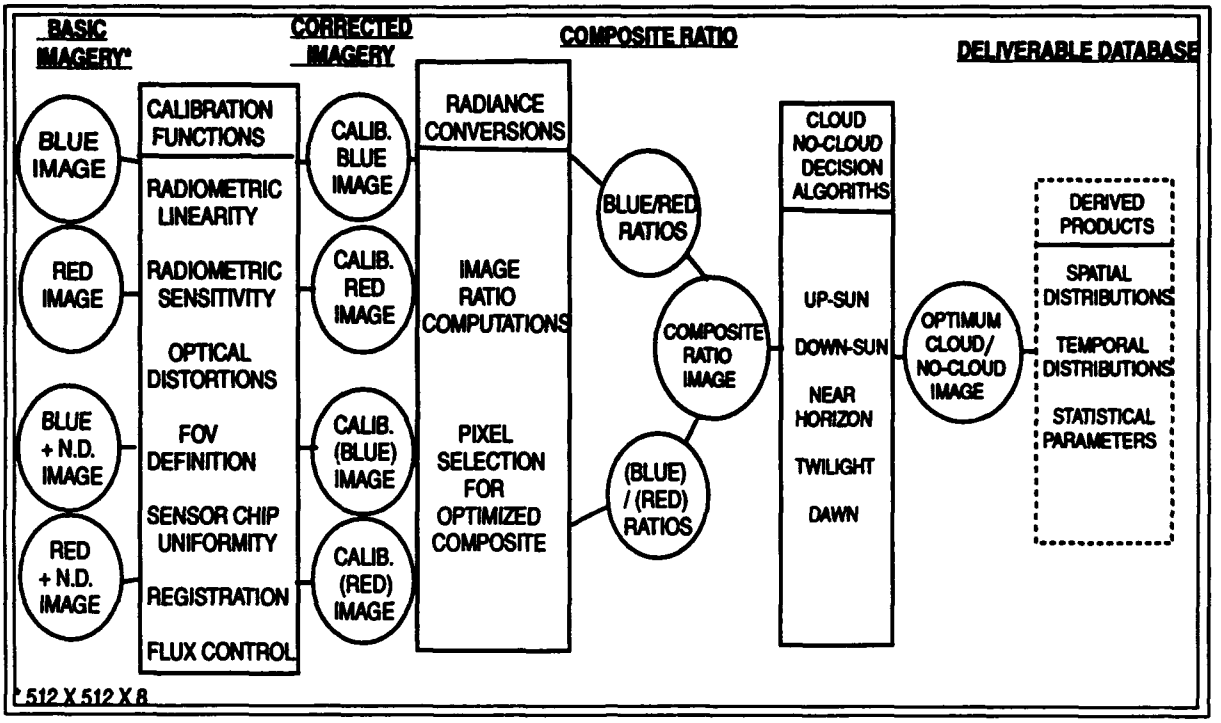

Figure A.3. WSI basic image processing flow chart.

to no data, clear, thin-cloud, thick-cloud, and off-scale bright (25). The result is a transformed image consisting of pixels at the five grey values described above.

Appendix B. *WSI Database Manipulation*

The purpose of this appendix is to provide an audit trail which traces the image data from object space to the working dataset used for our analysis.

Object Space to Image Space

The relationship between object space and image space is hyperbolic since the WSI system uses a fish-eye lens. The actual position in object space (θ, ϕ)[1] is related to the image position (x, y) by the following formulas:

Zenith angle:

$$\theta = 81° \sqrt{1.25(x-255)^2 + (y-240)^2}/230$$

Azimuth angle:

$$\phi = \arctan(1.25(x-255)/(y-240))$$

In these equations, azimuth is given relative to true north (25).

Values for each pixel were saved at five discrete levels according to the format described in Appendix C.

Where,

0 = No data

100 = Clear

150 = Thin Cloud

[1] *θ is the zenith angle, and ϕ is the azimuth angle.*

200 = Thick Cloud

250 = Off-scale bright

Scripps Institute archives the image data in one-minute and ten-minute formats. We used data extracted from the one-minute formats for our analysis.

Reduction of Image Dataset

Jeff Yepez from the Geophysics Directorate at Phillips Laboratory wrote a software routine which extracted the 33 data points which correspond to the Lund and Shanklin template azimuths and elevations. This subset of the one-minute tapes included the header data off of the one-minute tape and the pixel values in the 0–250 format.

Dr. T.S. Kelso from AFIT refined the 33 data point subset by converting the 0–250 format to a more concise and usable 0–9 format.

Where,

0 = No data

1 = Clear

2 = Thin Cloud

3 = Thick Cloud

9 = Off-scale bright

A sample of the final 33-point subset used for our analysis is depicted below.

```
89-02-01 14:02  98   1   0   0   0 0 10000000 00100000 00000000 00000000
89-02-01 14:03  91   9   1   0   0 0 10100000 10110000 00000000 00000000
89-02-01 14:04  80  19   1   0   0 0 11111000 10111000 00000000 00000000
89-02-01 14:05  66  30   4   0   0 1 21111100 21211200 10100000 10200000
89-02-01 14:06  58  38   4   0   0 1 11111100 11111200 10000000 11100000

YY-MM-DD HH:MM  SS  CC  TT  KK  BB Z NNNNNNNN EEEEEEEE SSSSSSSS WWWWWWWW
```

Where,

$YY - MM - DD$ = Year-Month-Day

$HH : MM$ = Hour:Minute

SS = Percent of Total Image Missing

CC = Percent of Total Image Clear

TT = Percent of Total Image Thin

KK = Percent of Total Image Thick

BB = Percentage of Total Image Off-Scale Bright

Z = Value at Zenith

$NNNNNNNN$ = Values of the eight observations on the North azimuth

$EEEEEEEE$ = Values of the eight observations on the East azimuth

$SSSSSSSS$ = Values of the eight observations on the South azimuth

$WWWWWWWW$ = Values of the eight observations on the West azimuth

All data used in our analysis (Columbia, Missouri and Kirtland AFB, New Mexico from February 1989 to March 1990) was converted from the one-minute data tape format from Scripps institute to the 33-point subset described above.

Appendix C. *Format of Ratio Tapes*

Technical Memorandum[1]

To: R. W. Johnson (13 Sep 90, AV90-123t)

From: J. E. Shields

Subject: Format of Ratio Tapes; One and Ten-Minute

This memo documents the format of WSI cloud ratio tapes. The format of both one and ten minute tapes is included in the description.

The ratio tapes are an intermediate step in the normal processing. As discussed in Tech Note 221, a variety of calibration functions are applied to the raw field data, which are then ratioed to create ratio tapes. These ratio tapes are then processed through a cloud decision program to yield the cloud decision tapes.

1. Contents of the Tapes

   A typical ratio tape contains either the one minute ratio images for a week or the ten minute ratio images for a week. Tapes may contain smaller amounts of data, depending on the contents of the field tape processed to generate the ratio tape.

   In the field, images are normally acquired at one minute intervals, for a total of 12 hours each day, centered on local apparent noon. Every 10 minutes, images are saved at full resolution, in 512 × 480 format. The 1-minute raw data images consist of 33 rows and 33 columns, each saved at full resolution for that line, thus creating a "screen" pattern as illustrated in Figure C.1.

   The ratio tape created by the Taprat program normally contains an EOF (End Of File) mark at the beginning of the tape, and at the end of each data day. Copy tapes of the ratio

[1] This appendix is a reproduction of Technical Memorandum AV90-123t prepared by J.E. Shields (24).

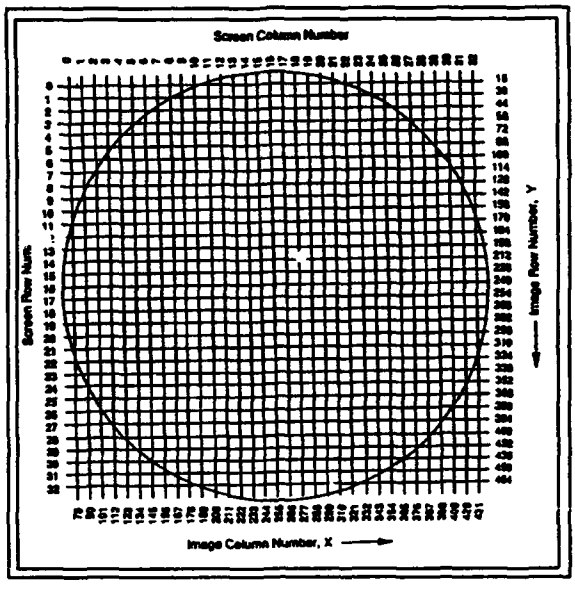

Figure C.1. Extracted one-minute image format.

will have one or more additional EOFs at the start and end of the tape. (If the copy was stopped and restarted in the middle, there will be additional EOF marks in the middle also.)

2. Data Structure for 10-minute Images

Each 10-minute image set consists of 1 logical block which contains a DOS-format header, followed by 480 logical blocks containing image data. Each logical block is 1024 bytes long.

In the DOS header block, the first 4 bytes (bytes 0-3) indicate the file size in bytes, which is 491,520 bytes. Byte 4 is the beginning of a string of ASCII characters which read "START FILENAME=filename". The filename is of the form

RUYYMMDD.TNN

where,

R = "Ratio"

U = Unit number, for example, unit 9 is the portable unit

YY = Year

MM = Month

DD = Day

T = Type, for example, 0 for One, T for Ten

NN = Number, for example, image number starting from the start of the day.

Thus for Dec 1, 1989, Unit 9, ten minute, the first image should have the name R9891201.T01. The remainder of the DOS header line consists of information pertaining to file size and creation date, stored in ASCII format.

On these 10 minute ratio tapes, the DOS header block is followed by 480 image blocks, each 1024 bytes long. These blocks contain the ratio data, plus the simultaneous red image acquired with Spectral Filter 4, and embedded header data. If these 480 lines are extracted from Exabyte and saved in rows 0 through 479 of an imaging board, they should appear in the format illustrated in Figure C.2. That is, the first half of each row contains the ratio information, and the second half contains the red image. When all 480 rows are read in, the result is a 512 × 480 ratio image in the first quadrant of the board, and a 512 × 480 red radiance image in the second quadrant of the board.

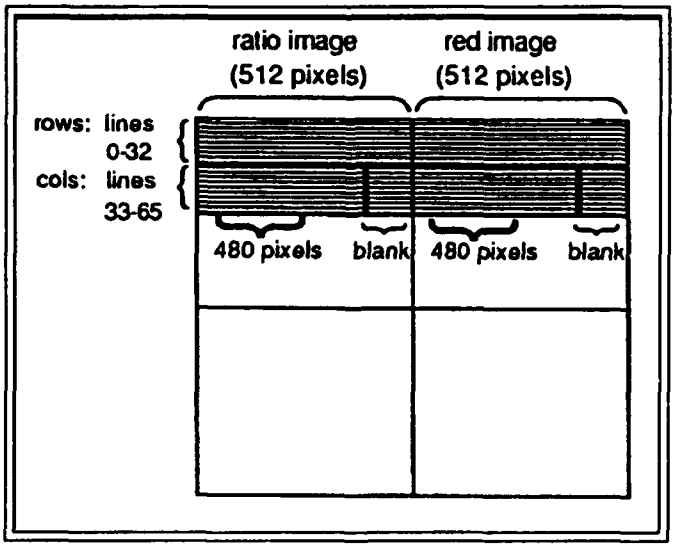

Figure C.2. Ten-minute ratio data extracted from Exabyte.

Portions of the first line of data also include embedded header information which documents the time and date, instrument status, and data quality information. The format of the embedded headers is documented in Memo AV90-042t (New Ratio and One-Minute Cloud Decision Header Format).

3. Data Structure for One Minute Images

Each 1-minute image set consists of 1 logical block containing a DOS header, followed by 66 logical blocks containing image data. Each logical block is 1024 bytes long. The DOS header information is in the same format as described above.

The contents of the 66 blocks of image data are illustrated in Figure C.3. If the 66 blocks are read directly to an FG100 board, the result would be as shown in Figure C.3. The first 512 pixels contain the rows from the Spectral 4 red image; and the last 32 pixels are blank.

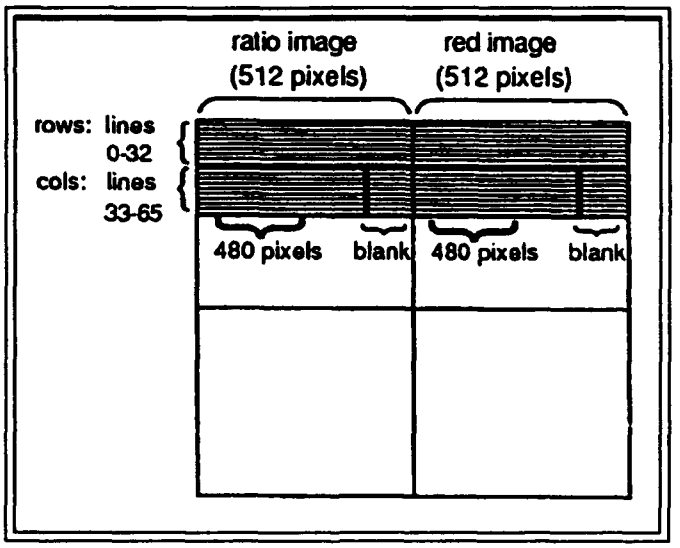

Figure C.3. One-minute ratio data extracted from Exabyte.

Like the 10-minute ratio images, the 1-minute images have header information embedded in the first block of image information, in the format described in Memo AV90-042t. Note that this header information is not the same as that contained in the DOS header. That is, the first logical block contains DOS header information, allowing us to use various utilities for

accessing files on Exabyte tape. The first block of image information includes embedded header information, used in processing and interpreting the data.

4. The Ratio Image

The ratio image is 8 bits deep, i.e. in 0 to 255 format. These should be considered relative ratios. That is, within the limits of measurement accuracy and signal to noise limitations, a value of 128 represents a ratio twice as high as the ratio corresponding to a value of 64. Until the calibrations are finalized, however, we cannot state what real ratio a value of 128 corresponds to.

Ratio values of 0 are used to indicate "no data." Any ratios exceeding the value of 239 are replaced by the value 239. The value 240 corresponds to points for which the blue or red radiance (not ratio) were off-scale bright. The values from 241 through 255 are reserved for graphics.

The relation between pixel location in image space and direction in object space is discussed in Memo AV90-041t. These images have been corrected to the standard pixel locations; that is, they have been adjusted for image size and location. Corrections for specific differences in lens geometry unique to a specific lens are applied in later processing steps.

5. The Radiance Image

As noted earlier, the Spectral 4 red image is also included on the ratio tapes. This image is included simply to allow us to make a visual assessment of the sky conditions. The image is not quite identical to the raw field image from Spectral 4. It has had the linearity correction applied; this calibration adjusts for the sensor chip signal versus radiance non-linearity. No other calibrations have been applied to the saved red image, since it is intended for visual assessment only.

JES:cr

Appendix D. *Contingency Table Outputs from SAS*

ELEVATION = 10 DEGREES
CUMULATIVE, FEBRUARY 89 TO JANUARY 90

TABLE OF AZIMUTH BY OBSERV

AZIMUTH / OBSERV	CLEAR	THICK	THIN	Total
EAST	124507	83567	23213	231287
	13.66	9.17	2.55	25.38
	53.83	36.13	10.04	
	25.70	24.45	27.32	
NORTH	120658	89357	21164	231179
	13.24	9.81	2.32	25.37
	52.19	38.65	9.15	
	24.91	26.14	24.91	
SOUTH	115018	83795	19040	217853
	12.62	9.20	2.09	23.91
	52.80	38.46	8.74	
	23.74	24.51	22.41	
WEST	124227	85101	21560	230888
	13.63	9.34	2.37	25.34
	53.80	36.86	9.34	
	25.65	24.90	25.37	
Total	484410	341820	84977	911207
	53.16	37.51	9.33	100.00

STATISTICS FOR TABLE OF AZIMUTH BY OBSERV

Statistic	DF	Value	Prob
Chi-Square	6	573.227	0.000

Sample Size = 911207

ELEVATION = 30 DEGREES
FEBRUARY 1989

TABLE OF AZIMUTH BY OBSERV

AZIMUTH / OBSERV	CLEAR	THICK	THIN	Total
EAST	5791	6311	758	12860
	11.28	12.30	1.48	25.06
	45.03	49.07	5.89	
	24.61	25.50	24.94	
NORTH	6061	6039	764	12864
	11.81	11.77	1.49	25.07
	47.12	46.94	5.94	
	25.76	24.40	25.14	
SOUTH	5760	6300	672	12732
	11.22	12.28	1.31	24.81
	45.24	49.48	5.28	
	24.48	25.45	22.11	
WEST	5915	6103	845	12863
	11.53	11.89	1.65	25.06
	45.98	47.45	6.57	
	25.14	24.66	27.81	
Total	23527	24753	3039	51319
	45.84	48.23	5.92	100.00

STATISTICS FOR TABLE OF AZIMUTH BY OBSERV

Statistic	DF	Value	Prob
Chi-Square	6	37.579	0.000

Sample Size = 51319

ELEVATION = 30 DEGREES
MARCH 1989

TABLE OF AZIMUTH BY OBSERV

AZIMUTH / OBSERV	CLEAR	THICK	THIN	Total
EAST	10478	9763	1471	21712
	12.49	11.63	1.75	25.87
	48.26	44.97	6.78	
	25.26	26.52	26.19	
NORTH	11173	9182	1400	21755
	13.31	10.94	1.67	25.92
	51.36	42.21	6.44	
	26.93	24.94	24.92	
SOUTH	9114	8201	1450	18765
	10.86	9.77	1.73	22.36
	48.57	43.70	7.73	
	21.97	22.27	25.81	
WEST	10719	9674	1296	21689
	12.77	11.53	1.54	25.84
	49.42	44.60	5.98	
	25.84	26.27	23.07	
Total	41484	36820	5617	83921
	49.43	43.87	6.69	100.00

STATISTICS FOR TABLE OF AZIMUTH BY OBSERV

Statistic	DF	Value	Prob
Chi-Square	6	96.691	0.000

Sample Size = 83921

ELEVATION = 30 DEGREES
APRIL 1989

TABLE OF AZIMUTH BY OBSERV

AZIMUTH / OBSERV	CLEAR	THICK	THIN	Total
EAST	8968	8002	2030	19000
	12.44	11.10	2.82	26.36
	47.20	42.12	10.68	
	24.96	27.48	28.88	
NORTH	10008	7330	1678	19016
	13.88	10.17	2.33	26.38
	52.63	38.55	8.82	
	27.85	25.17	23.88	
SOUTH	7486	6131	1621	15238
	10.39	8.51	2.25	21.14
	49.13	40.23	10.64	
	20.83	21.05	23.06	
WEST	9469	7657	1699	18825
	13.14	10.62	2.36	26.12
	50.30	40.67	9.03	
	26.35	26.29	24.17	
Total	35931	29120	7028	72079
	49.85	40.40	9.75	100.00

STATISTICS FOR TABLE OF AZIMUTH BY OBSERV

Statistic	DF	Value	Prob
Chi-Square	6	145.253	0.000

Sample Size = 72079

ELEVATION = 30 DEGREES
MAY 1989

TABLE OF AZIMUTH BY OBSERV

AZIMUTH / OBSERV	CLEAR	THICK	THIN	Total
EAST	4769	6092	1171	12032
Percent	9.54	12.19	2.34	24.07
Row Pct	39.64	50.63	9.73	
Col Pct	24.18	23.15	29.68	
NORTH	5729	7599	1091	14419
Percent	11.46	15.20	2.18	28.85
Row Pct	39.73	52.70	7.57	
Col Pct	29.04	28.88	27.65	
SOUTH	4605	6127	878	11610
Percent	9.21	12.26	1.76	23.23
Row Pct	39.66	52.77	7.56	
Col Pct	23.34	23.29	22.25	
WEST	4623	6494	806	11923
Percent	9.25	12.99	1.61	23.85
Row Pct	38.77	54.47	6.76	
Col Pct	23.44	24.68	20.43	
Total	19726	26312	3946	49984
	39.46	52.64	7.89	100.00

STATISTICS FOR TABLE OF AZIMUTH BY OBSERV

Statistic	DF	Value	Prob
Chi-Square	6	93.238	0.000

Sample Size = 49984

ELEVATION = 30 DEGREES
JUNE 1989

TABLE OF AZIMUTH BY OBSERV

AZIMUTH OBSERV

Frequency / Percent / Row Pct / Col Pct	CLEAR	THICK	THIN	Total
EAST	2313	1448	900	4661
	11.65	7.29	4.53	23.48
	49.62	31.07	19.31	
	22.24	22.45	29.97	
NORTH	2973	1995	896	5864
	14.97	10.05	4.51	29.54
	50.70	34.02	15.28	
	28.59	30.93	29.84	
SOUTH	2549	1510	684	4743
	12.84	7.61	3.45	23.89
	53.74	31.84	14.42	
	24.51	23.41	22.78	
WEST	2565	1498	523	4586
	12.92	7.55	2.63	23.10
	55.93	32.66	11.40	
	24.66	23.22	17.42	
Total	10400	6451	3003	19854
	52.38	32.49	15.13	100.00

STATISTICS FOR TABLE OF AZIMUTH BY OBSERV

Statistic	DF	Value	Prob
Chi-Square	6	128.014	0.000

Sample Size = 19854

ELEVATION = 30 DEGREES
JULY 1989

TABLE OF AZIMUTH BY OBSERV

AZIMUTH	OBSERV			
Frequency Percent Row Pct Col Pct	CLEAR	THICK	THIN	Total
EAST	6062 8.54 36.67 25.59	8561 12.06 51.79 22.68	1907 2.69 11.54 19.91	16530 23.28
NORTH	6645 9.36 31.58 28.05	11233 15.82 53.38 29.76	3166 4.46 15.04 33.05	21044 29.64
SOUTH	5925 8.34 34.70 25.01	8712 12.27 51.03 23.08	2436 3.43 14.27 25.43	17073 24.04
WEST	5059 7.12 30.92 21.35	9233 13.00 56.43 24.47	2070 2.92 12.65 21.61	16362 23.04
Total	23691 33.36	37739 53.15	9579 13.49	71009 100.00

STATISTICS FOR TABLE OF AZIMUTH BY OBSERV

Statistic	DF	Value	Prob
Chi-Square	6	267.079	0.000

Sample Size = 71009

ELEVATION = 30 DEGREES
AUGUST 1989

TABLE OF AZIMUTH BY OBSERV

AZIMUTH / OBSERV	CLEAR	THICK	THIN	Total
EAST	9472	7748	2293	19513
	12.24	10.01	2.96	25.21
	48.54	39.71	11.75	
	24.84	24.93	27.96	
NORTH	11020	8189	2062	21271
	14.24	10.58	2.66	27.48
	51.81	38.50	9.69	
	28.90	26.35	25.14	
SOUTH	8842	6554	1789	17185
	11.42	8.47	2.31	22.20
	51.45	38.14	10.41	
	23.19	21.09	21.81	
WEST	8797	8590	2058	19445
	11.36	11.10	2.66	25.12
	45.24	44.18	10.58	
	23.07	27.64	25.09	
Total	38131	31081	8202	77414
	49.26	40.15	10.59	100.00

STATISTICS FOR TABLE OF AZIMUTH BY OBSERV

Statistic	DF	Value	Prob
Chi-Square	6	263.321	0.000

Sample Size = 77414

ELEVATION = 30 DEGREES
SEPTEMBER 1989

TABLE OF AZIMUTH BY OBSERV

AZIMUTH	OBSERV			
Frequency / Percent / Row Pct / Col Pct	CLEAR	THICK	THIN	Total
EAST	11429	7497	1586	20512
	14.62	9.59	2.03	26.25
	55.72	36.55	7.73	
	25.99	25.61	32.29	
NORTH	11899	7597	1067	20563
	15.23	9.72	1.37	26.31
	57.87	36.94	5.19	
	27.06	25.96	21.72	
SOUTH	9449	6106	1010	16565
	12.09	7.81	1.29	21.20
	57.04	36.86	6.10	
	21.49	20.86	20.56	
WEST	11196	8069	1249	20514
	14.33	10.32	1.60	26.25
	54.58	39.33	6.09	
	25.46	27.57	25.43	
Total	43973	29269	4912	78154
	56.26	37.45	6.29	100.00

STATISTICS FOR TABLE OF AZIMUTH BY OBSERV

Statistic	DF	Value	Prob
Chi-Square	6	159.277	0.000

Sample Size = 78154

ELEVATION = 30 DEGREES
OCTOBER 1989

TABLE OF AZIMUTH BY OBSERV

AZIMUTH / OBSERV	CLEAR	THICK	THIN	Total
EAST	13632	5405	1420	20457
	16.92	6.71	1.76	25.40
	66.64	26.42	6.94	
	26.16	23.04	28.49	
NORTH	13558	5675	1190	20423
	16.83	7.04	1.48	25.35
	66.39	27.79	5.83	
	26.02	24.19	23.88	
SOUTH	12499	5940	850	19289
	15.52	7.37	1.06	23.95
	64.80	30.79	4.41	
	23.99	25.32	17.05	
WEST	12418	6443	1524	20385
	15.42	8.00	1.89	25.31
	60.92	31.61	7.48	
	23.83	27.46	30.58	
Total	52107	23463	4984	80554
	64.69	29.13	6.19	100.00

STATISTICS FOR TABLE OF AZIMUTH BY OBSERV

Statistic	DF	Value	Prob
Chi-Square	6	368.079	0.000

Sample Size = 80554

ELEVATION = 30 DEGREES
NOVEMBER 1989

TABLE OF AZIMUTH BY OBSERV

AZIMUTH / OBSERV	CLEAR	THICK	THIN	Total
EAST	13521	3598	1023	18142
	18.66	4.96	1.41	25.03
	74.53	19.83	5.64	
	25.85	22.56	24.25	
NORTH	13455	3775	907	18137
	18.56	5.21	1.25	25.02
	74.19	20.81	5.00	
	25.72	23.67	21.50	
SOUTH	12728	4322	1042	18092
	17.56	5.96	1.44	24.96
	70.35	23.89	5.76	
	24.33	27.10	24.70	
WEST	12605	4255	1246	18106
	17.39	5.87	1.72	24.98
	69.62	23.50	6.88	
	24.10	26.68	29.54	
Total	52309	15950	4218	72477
	72.17	22.01	5.82	100.00

STATISTICS FOR TABLE OF AZIMUTH BY OBSERV

Statistic	DF	Value	Prob
Chi-Square	6	204.126	0.000

Sample Size = 72477

ELEVATION = 30 DEGREES
DECEMBER 1989

TABLE OF AZIMUTH BY OBSERV

AZIMUTH / OBSERV	CLEAR	THICK	THIN	Total
EAST	2447	816	247	3510
	17.46	5.82	1.76	25.04
	69.72	23.25	7.04	
	26.92	22.20	19.71	
NORTH	2337	850	316	3503
	16.67	6.06	2.25	24.99
	66.71	24.26	9.02	
	25.71	23.13	25.22	
SOUTH	2169	986	350	3505
	15.47	7.03	2.50	25.01
	61.88	28.13	9.99	
	23.86	26.83	27.93	
WEST	2136	1023	340	3499
	15.24	7.30	2.43	24.96
	61.05	29.24	9.72	
	23.50	27.84	27.13	
Total	9089	3675	1253	14017
	64.84	26.22	8.94	100.00

STATISTICS FOR TABLE OF AZIMUTH BY OBSERV

Statistic	DF	Value	Prob
Chi-Square	6	82.116	0.000

Sample Size = 14017

ELEVATION = 30 DEGREES
JANUARY 1990

TABLE OF AZIMUTH BY OBSERV

AZIMUTH OBSERV

Frequency / Percent / Row Pct / Col Pct	CLEAR	THICK	THIN	Total
EAST	12175	2217	2279	16671
	18.27	3.33	3.42	25.02
	73.03	13.30	13.67	
	27.50	13.88	35.61	
NORTH	11201	4077	1386	16664
	16.81	6.12	2.08	25.01
	67.22	24.47	8.32	
	25.30	25.53	21.66	
SOUTH	10576	4755	1295	16626
	15.87	7.14	1.94	24.95
	63.61	28.60	7.79	
	23.89	29.78	20.24	
WEST	10315	4919	1439	16673
	15.48	7.38	2.16	25.02
	61.87	29.50	8.63	
	23.30	30.81	22.49	
Total	44267	15968	6399	66634
	66.43	23.96	9.60	100.00

STATISTICS FOR TABLE OF AZIMUTH BY OBSERV

Statistic	DF	Value	Prob
Chi-Square	6	1727.962	0.000

Sample Size = 66634

Appendix E. *Trend Analysis Plots*

The purpose of this appendix is to display the results of the trend analysis tests. The results for each direction are presented in two graphs.

The bottom graph displays the data points that represent the correlation by lags. As shown in Table E.1, the number of data points decreases for each successive lag because of the larger distances between neighbors. So, for lag = 1, there are seven data points; for lag = 2, six data points; for lag = 3, five data points; et cetera.

Table E.1. Matrix of data points per lag.

Lags	*Data Points*						
1	10-20	20-30	30-40	40-50	50-60	60-70	70-80
2	10-30	20-40	30-50	40-60	50-70	60-80	
3	10-40	20-50	30-60	40-70	50-80		
4	10-50	20-60	30-70	40-80			
5	10-60	20-70	30-50				
6	10-70	20-80					

In the top graph, a mean and standard deviation are calculated and plotted for each lag. For example, the seven data points for lag = 1 are averaged together and a mean and standard deviation are computed.

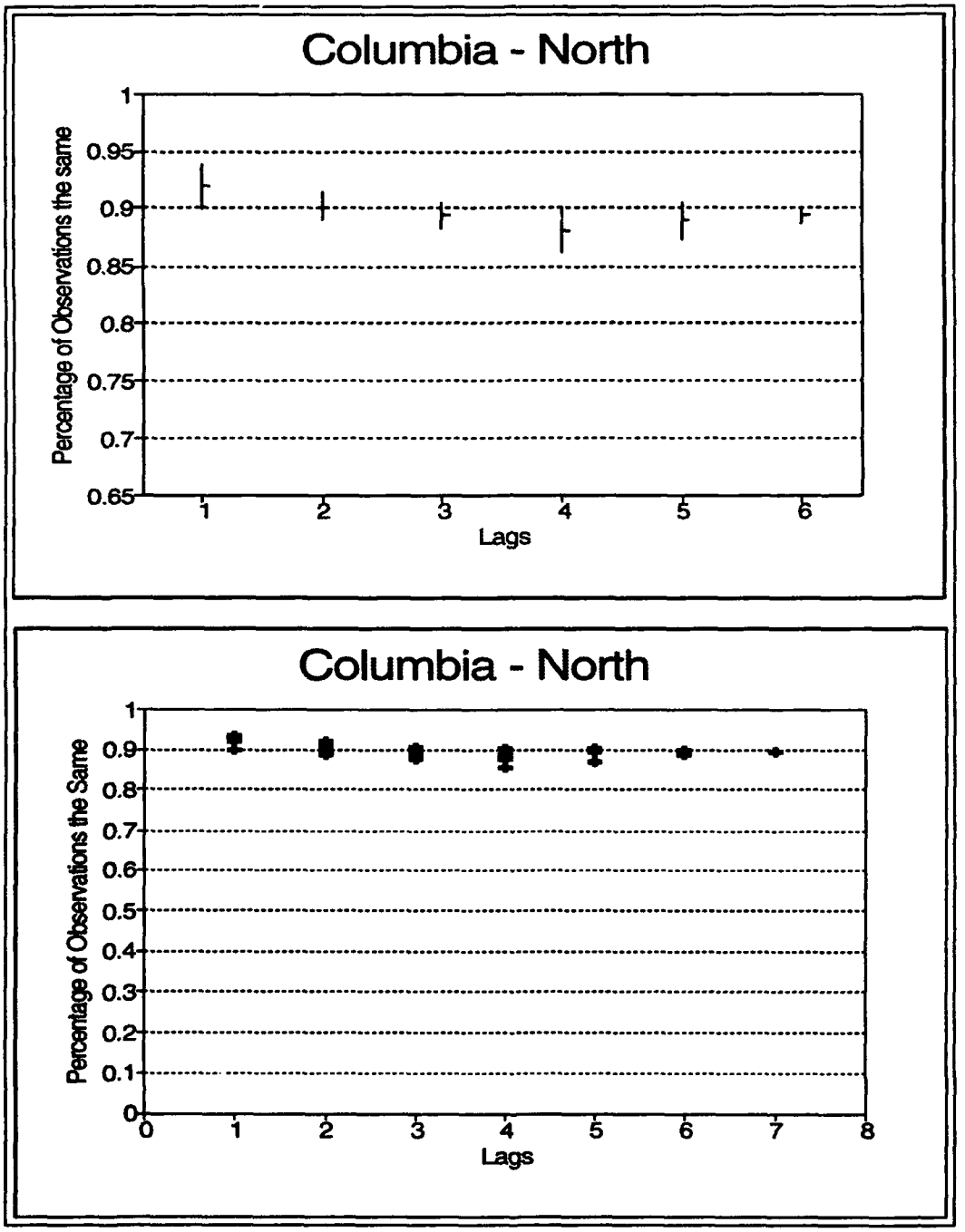

Figure E.1. Plots of correlation for k spatial intervals along the north axis. Bottom: plot of correlation data points. Top: plot of mean and one standard deviation confidence intervals.

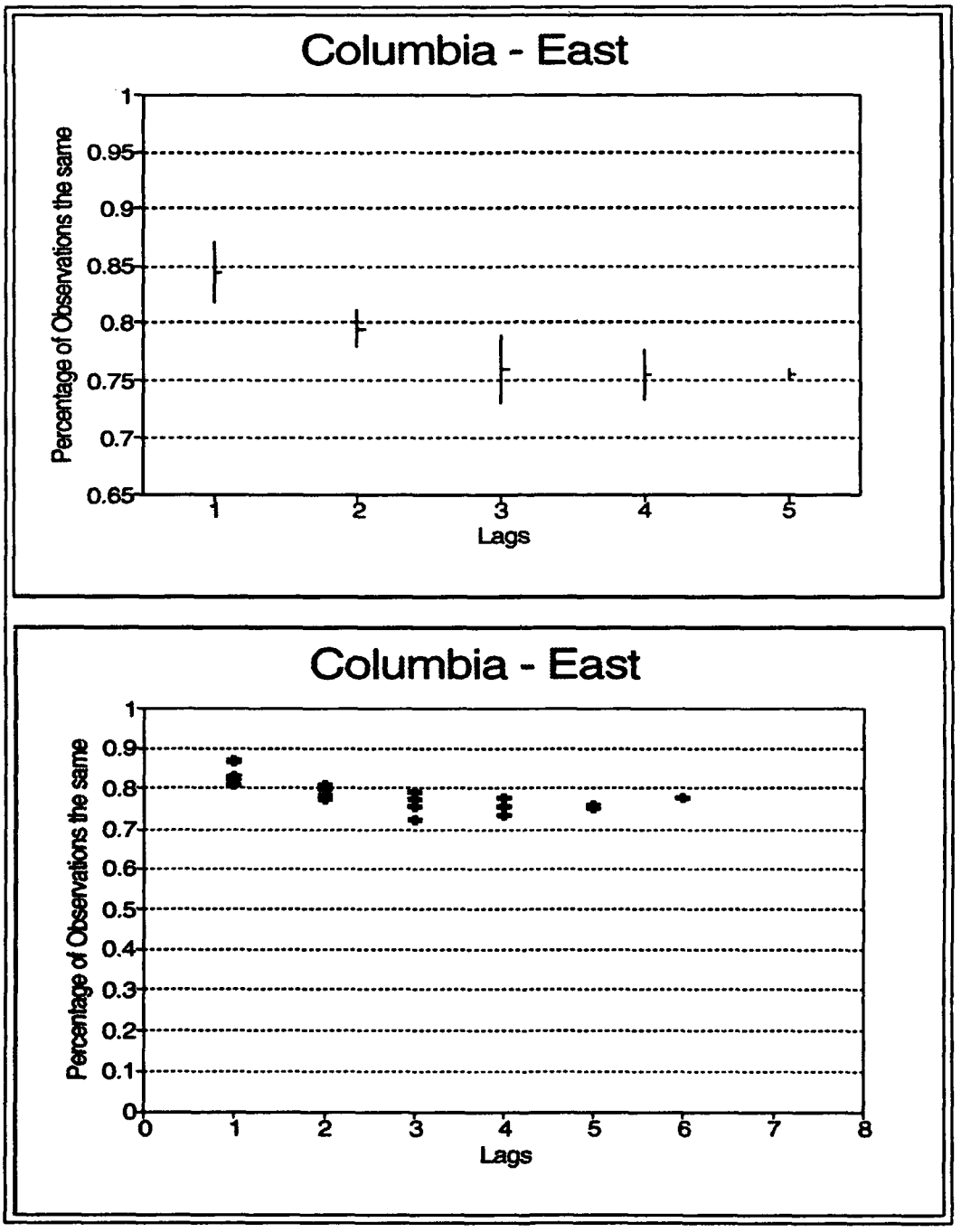

Figure E.2. Plots of correlation for k spatial intervals along the east axis. Bottom: plot of correlation data points. Top: plot of mean and one standard deviation confidence intervals.

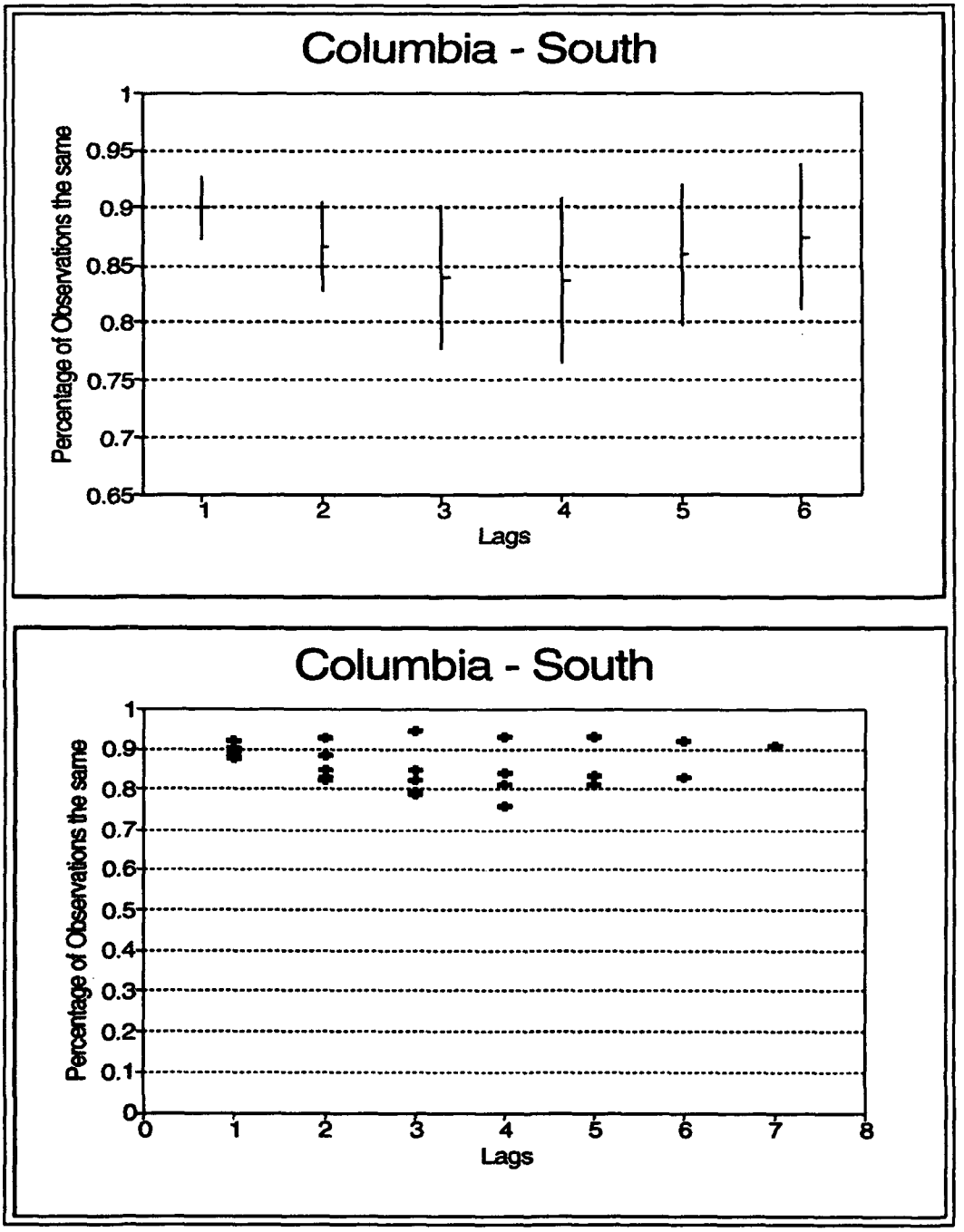

Figure E.3. Plots of correlation for k spatial intervals along the south axis. Bottom: plot of correlation data points. Top: plot of mean and one standard deviation confidence intervals.

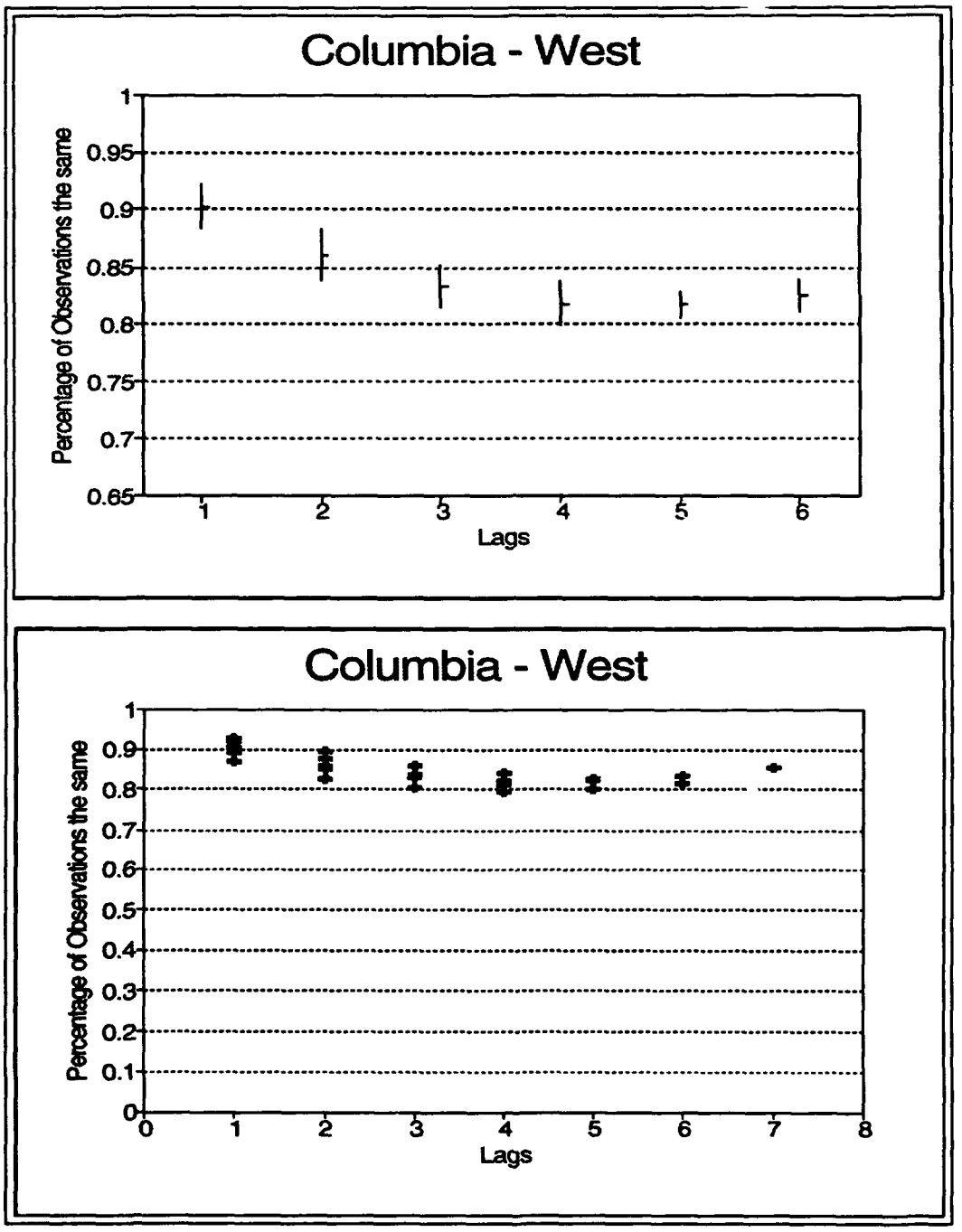

Figure E.4. Plots of correlation for k spatial intervals along the west axis. Bottom: plot of correlation data points. Top: plot of mean and one standard deviation confidence intervals.

Appendix F. *Time Series Analysis*

Identification of feasible models is accomplished by examining plots of the ACFs and PACFs of the original data and residuals. Cumulative probability plots and periodograms are also used to ensure the candidate models have accounted for all significant terms and have reduced the residuals to random (white) noise.

The first operation performed on the original time series was to difference the data to obtain a stationary series. Once a stationary time series was obtained, the ACFs and PACFs were calculated and plotted for the original data and residuals. Cumulative probability plots and periodograms were also produced for the residuals.

Plots of Original Time Series and Differenced Data

The purpose of this section is to display the original time-series data and the differenced data. A plot is presented for each cardinal direction. The raw time-series is displayed on top, the first differenced data is in the middle, and the second differenced data is on the bottom. Plots for the Columbia data are presented first and are followed by the plots of the Kirtland data.

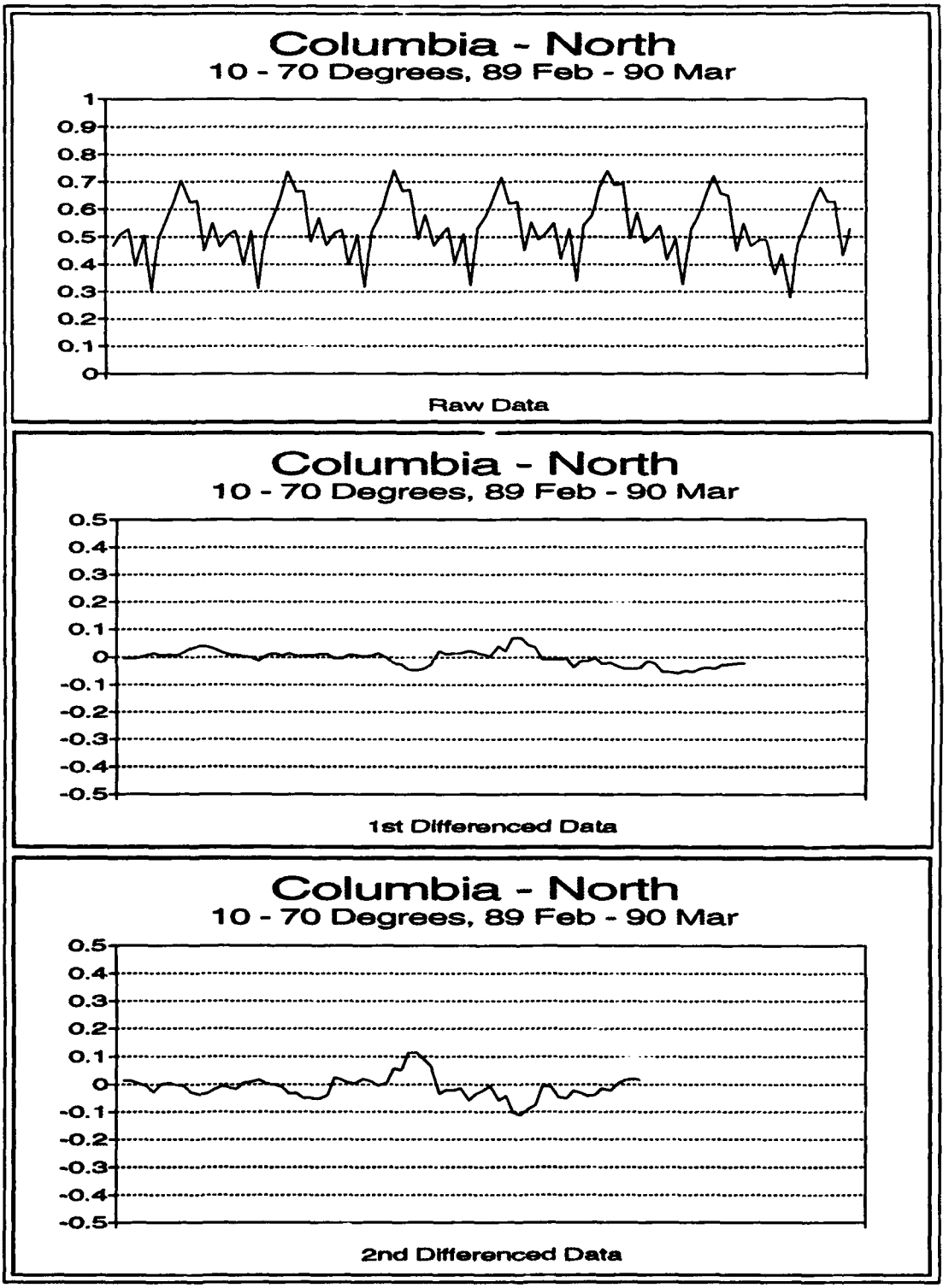

Figure F.1. Raw and differenced data for the north azimuth at Columbia, Missouri.

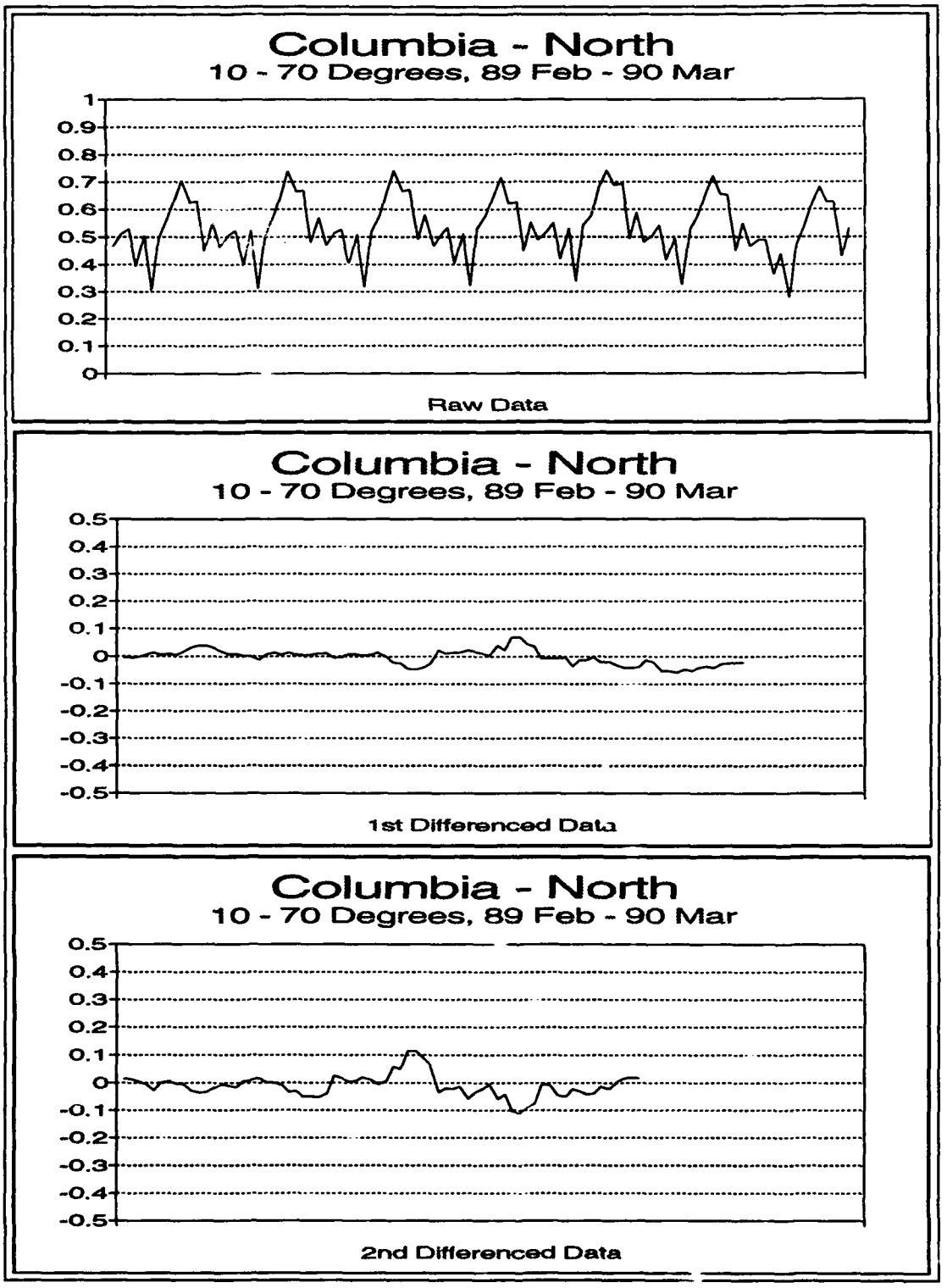

Figure F.1. Raw and differenced data for the north azimuth at Columbia, Missouri.

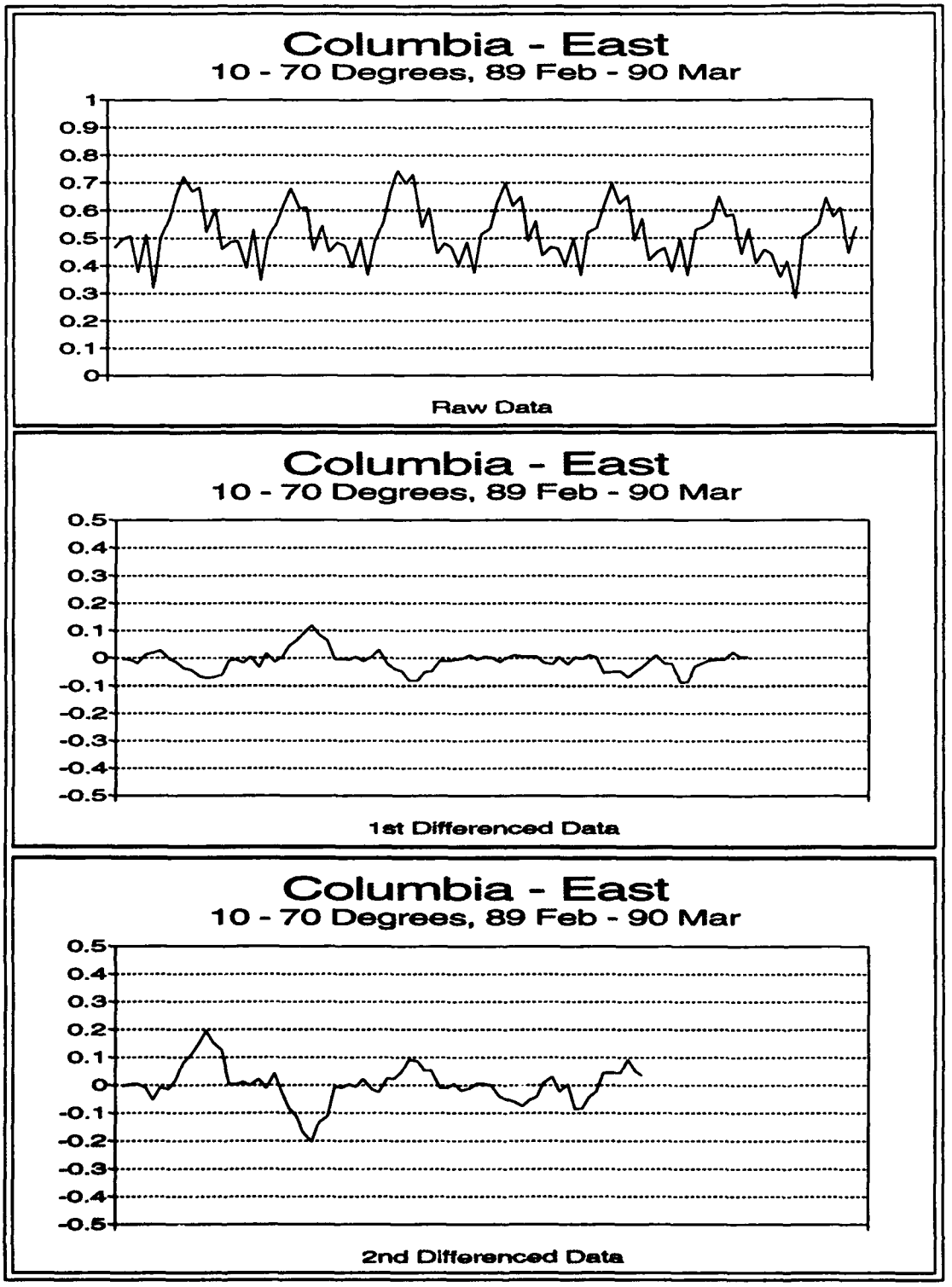

Figure F.2. Raw and differenced data for the east azimuth at Columbia, Missouri.

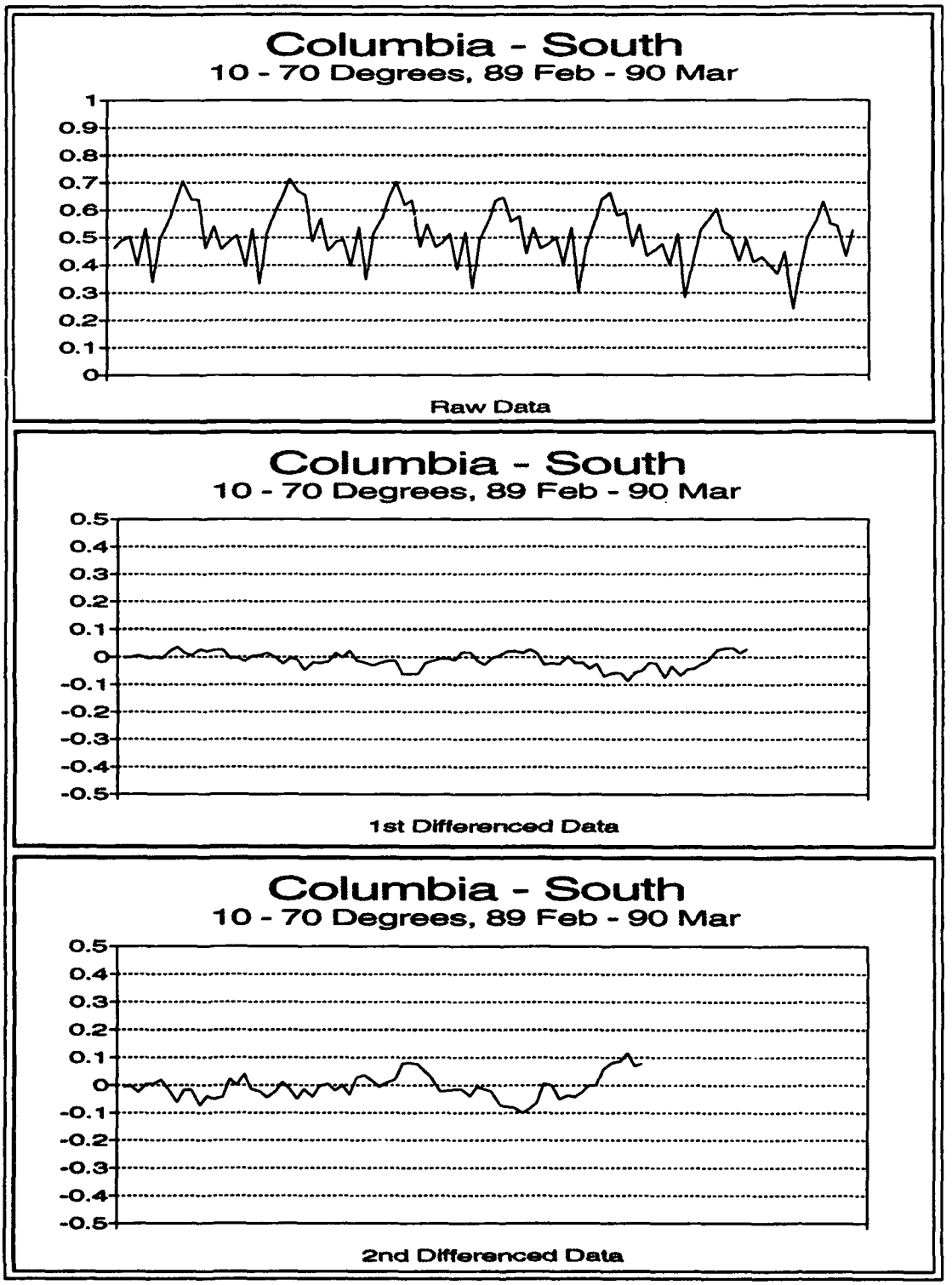

Figure F.3. Raw and differenced data for the south azimuth at Columbia, Missouri.

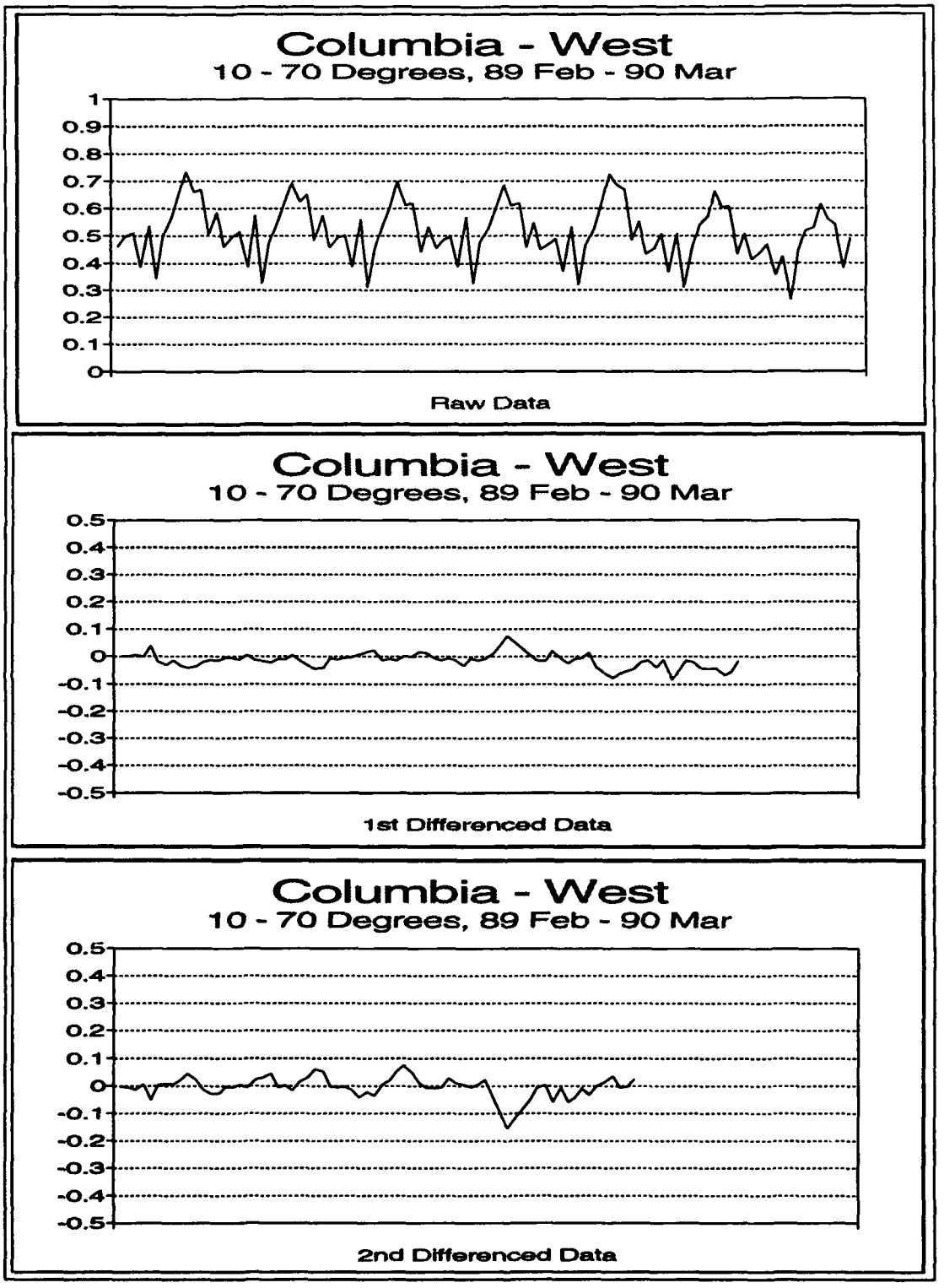

Figure F.4. Raw and differenced data for the west azimuth at Columbia, Missouri.

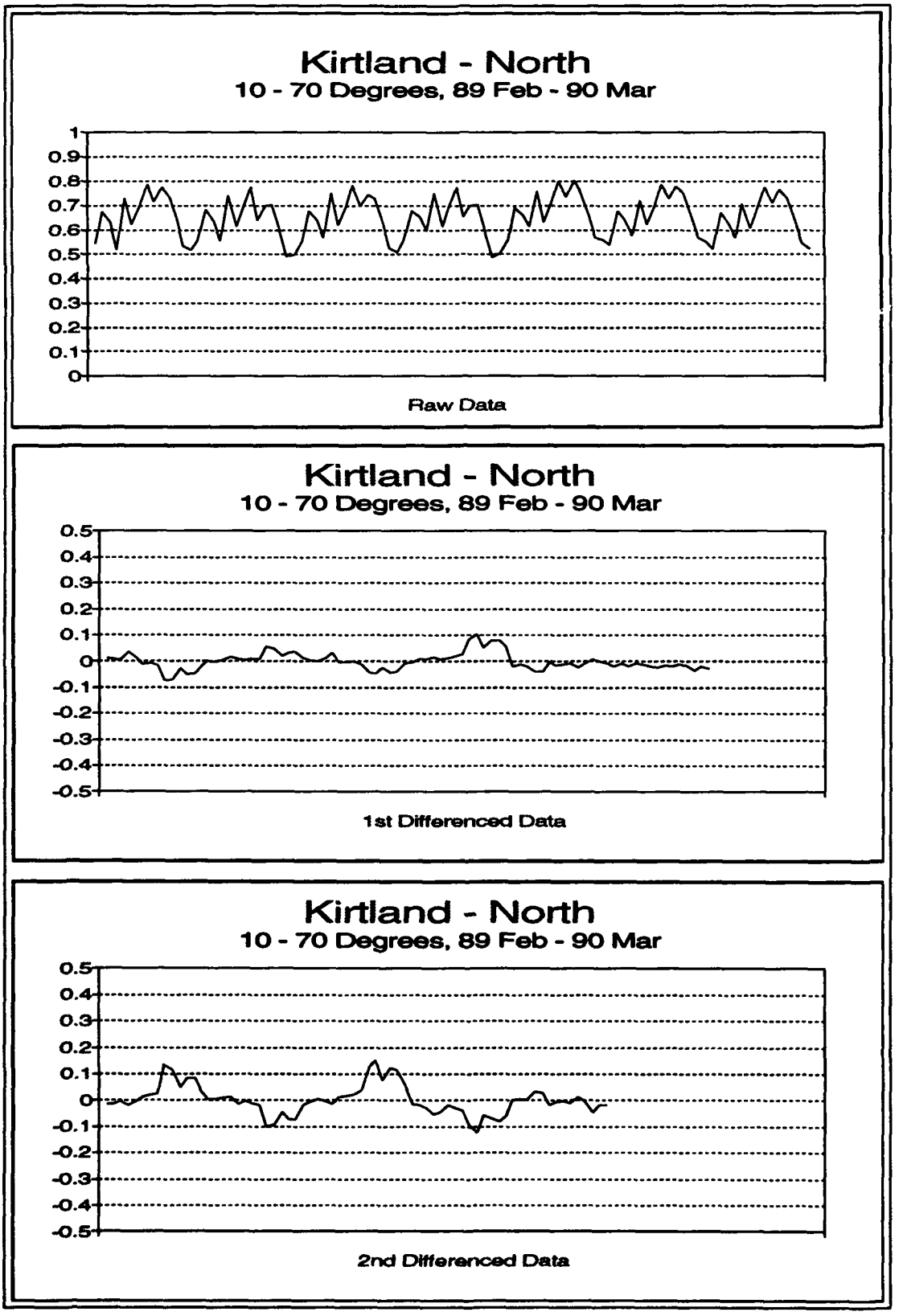

Figure F.5. Raw and differenced data for the north azimuth at Kirtland AFB, New Mexico.

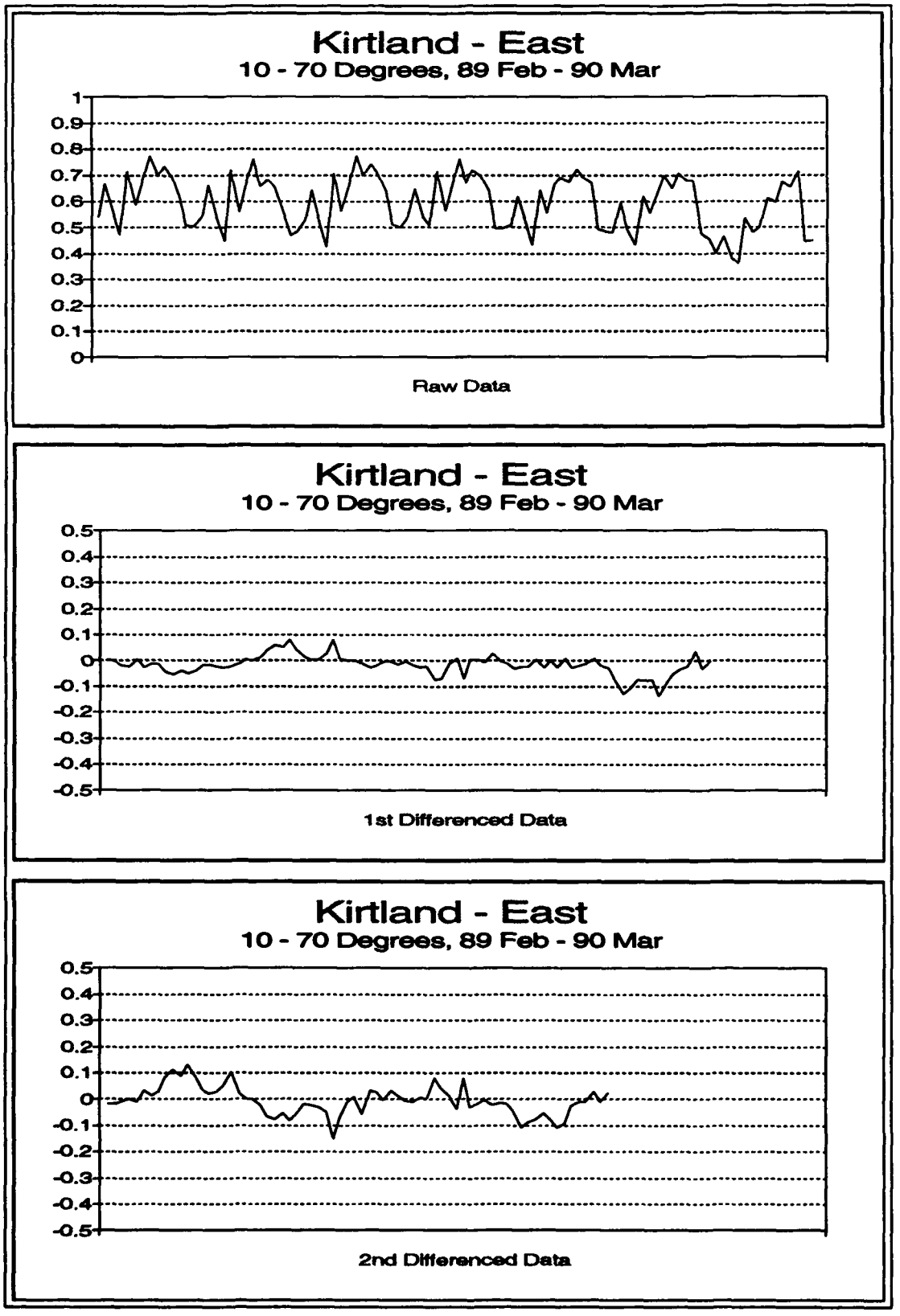

Figure F.6. Raw and differenced data for the east azimuth at Kirtland AFB, New Mexico.

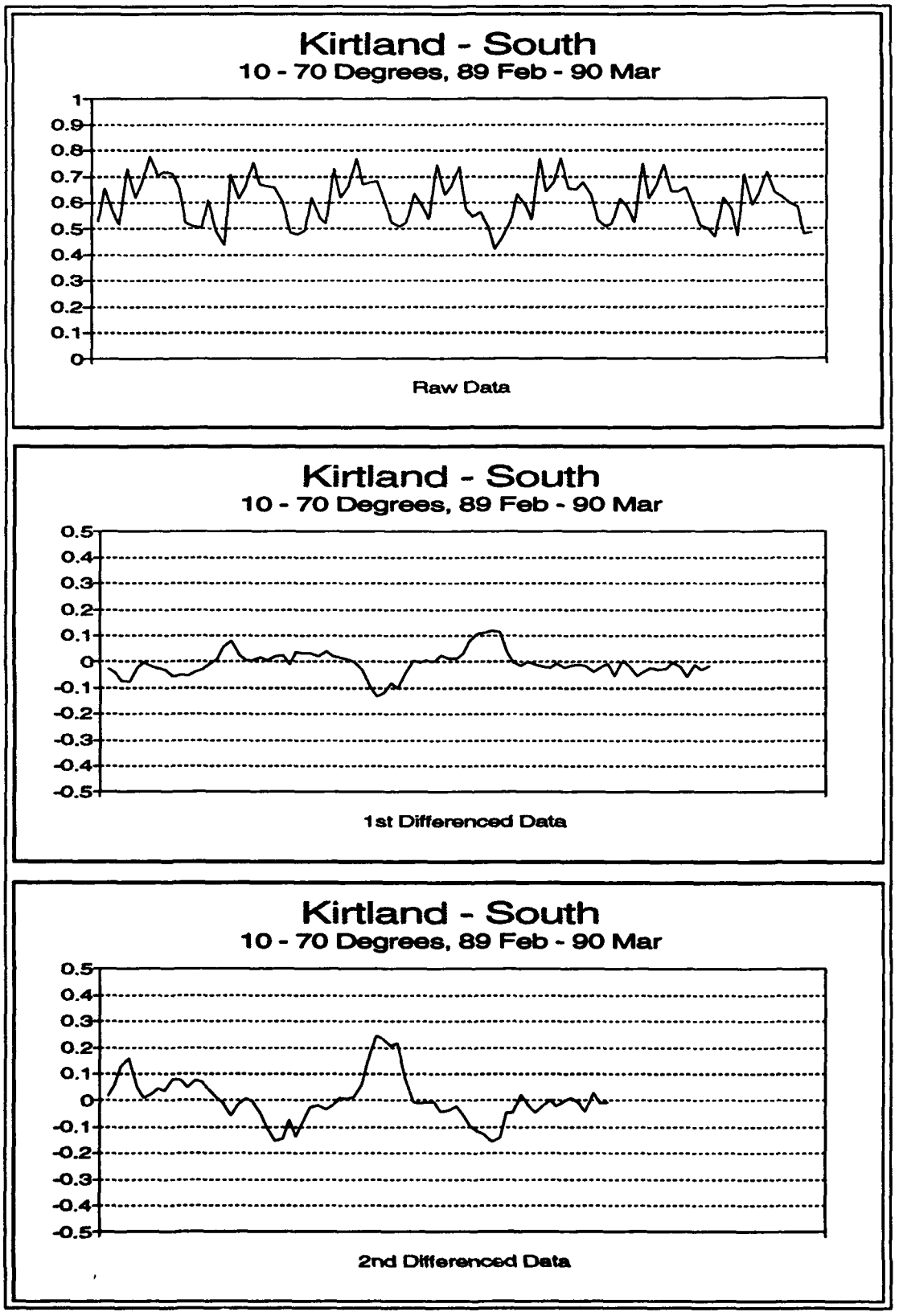

Figure F.7. Raw and differenced data for the south azimuth at Kirtland AFB, New Mexico.

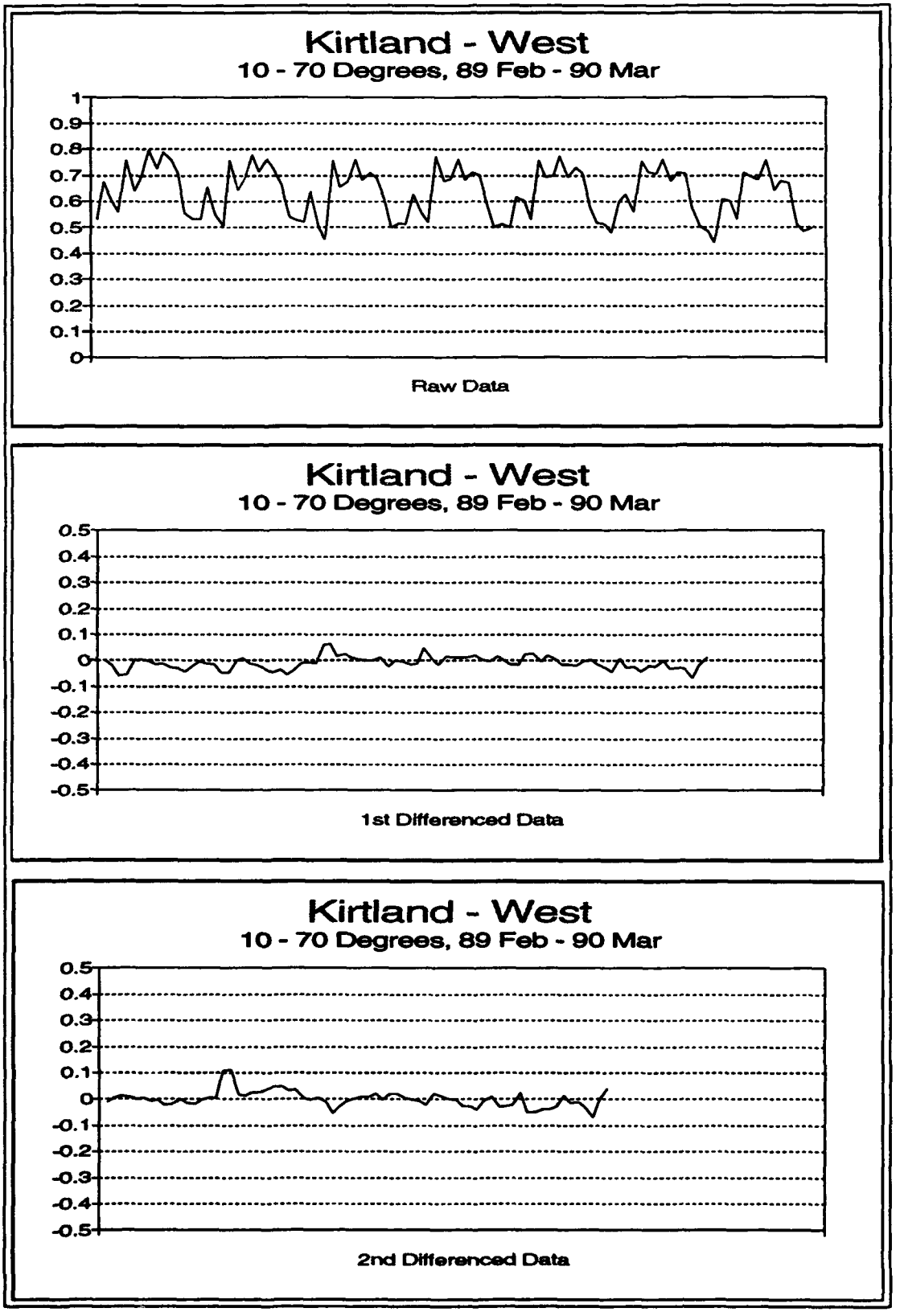

Figure F.8. Raw and differenced data for the west azimuth at Kirtland AFB, New Mexico.

Plots of ACFs and PACFs for the Original Data

The purpose of this section is to display the ACFs and PACFs of the original data. But first, a brief description of ACFs and PACFs is presented to explain the plots and how the Statgraphic routines calculate the functions.

Autocorrelation Coefficient Function (ACF). The ACF provides a means of testing for seasonal patterns in a time series. The Statgraphics ACF routine computes correlation coefficients between a time series variable and the values of that variable k time periods earlier. The Statgraphics ACF routine plots vertical bars representing the correlations for lags from one to k. The height of the bars represents the estimated correlation coefficients. On the ACF plots of residuals, dashed lines are plotted at plus and minus twice the standard errors for each coefficient. We use these boundaries to identify which correlation terms are significantly different from zero. The significant terms are used as a preliminary step to identify an appropriate parametric model[1].

The best estimate of the autocorrelation function, ρ_k, at lag k is r_k according to Box and Jenkins (3:32). The Statgraphics ACF routine uses Box and Jenkins' formulation of r_k to calculated the ACF estimates:

$$r_k = \frac{c_k}{c_0} = \frac{\sum_{t=1}^{N-k} (z_t - \bar{z})(z_{(t+k)} - \bar{z})}{\sum_{t=1}^{N} (z_t - \bar{z})^2}$$

where,

z_t = the observation at time t for $z_1, z_2, ..., z_n$ number of N observations.

$\bar{z}$ = the mean of the time series.

[1]This description of the plots and ACF application is paraphrased from the Statgraphics Reference Manual (27:A-13).

Partial Autocorrelation Coefficient Function (PACF). Analogous to the ACF, the PACF is useful in estimating the number of terms in an autoregressive model. The Statgraphics PACF routine plots a vertical bar for each coefficient. The height of the bar is proportional to the value of the coefficient. On the PACF plots of residuals, the PACF routine places dashed lines at $\pm 2 / \sqrt{n}$ to indicate which partial autocorrelations are significantly different from zero (27:P-23).

The Statgraphics PACF routine calculates the PACFs by solving the Yule-Walker equations:[2]

$$r_j = \hat{\phi}_{k1} r_{j-1} + \hat{\phi}_{k2} r_{j-2} + \ldots + \hat{\phi}_{k(k-1)} r_{j-k+1} + \hat{\phi}_{kk} r_{j-k}$$

where,

$j = 1, 2, \ldots, k$

$k = 1, 2, \ldots$

Accordingly, we used the Statgraphics ACF and PACF routines to calculate the first 20 lags of the original series for each site and direction. The plots of the ACFs and PACFs are presented according to site and direction. The Columbia plots for each direction are presented first and are followed by the Kirtland data.

[2]The Yule-Walker estimates of the successive autoregressive processes may only be employed if the values of the parameters are not too close to the non-stationary boundaries (3:65).

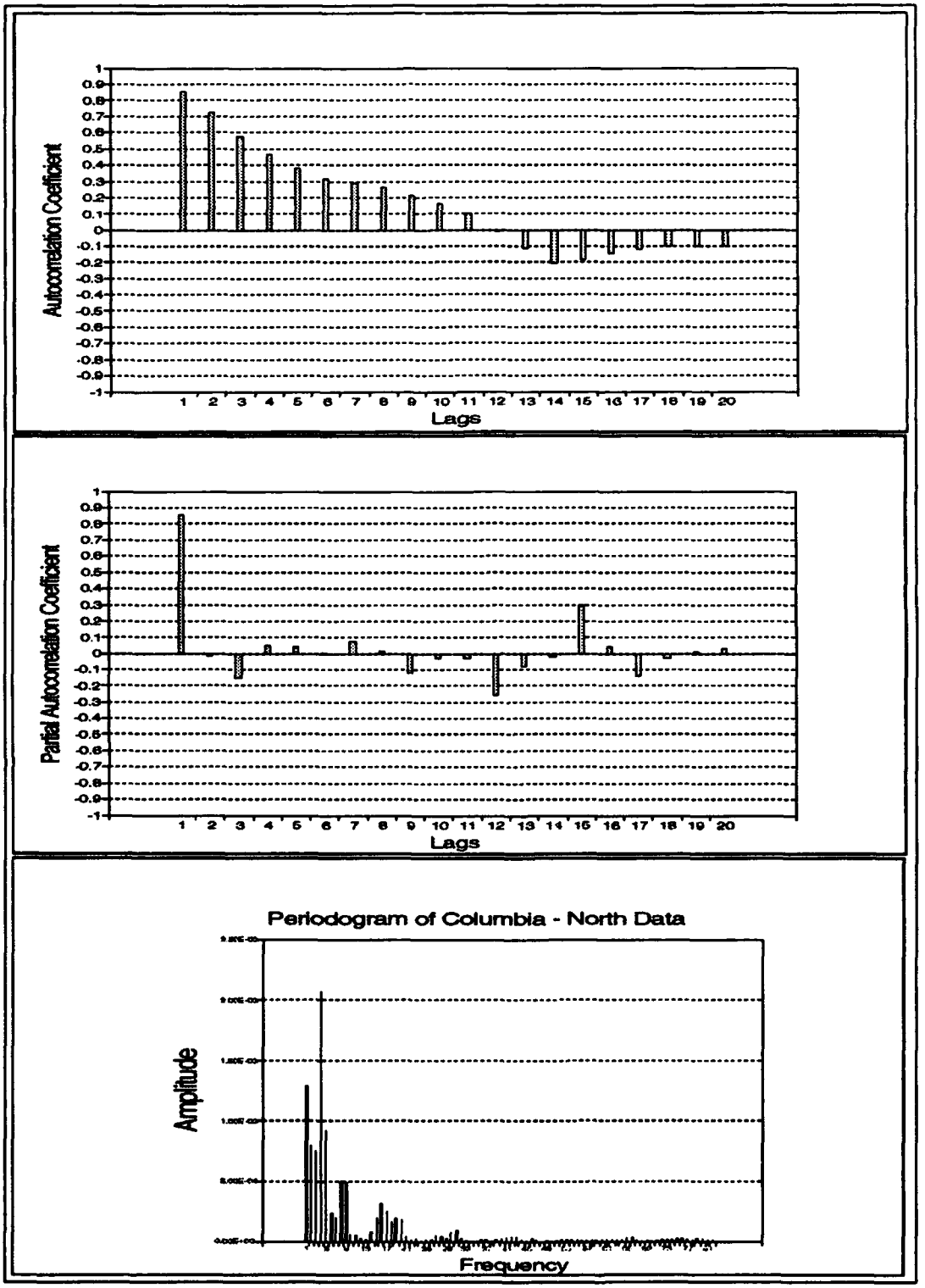

Figure F.9. ACF and PACF plots for the north azimuth at Columbia, Missouri.

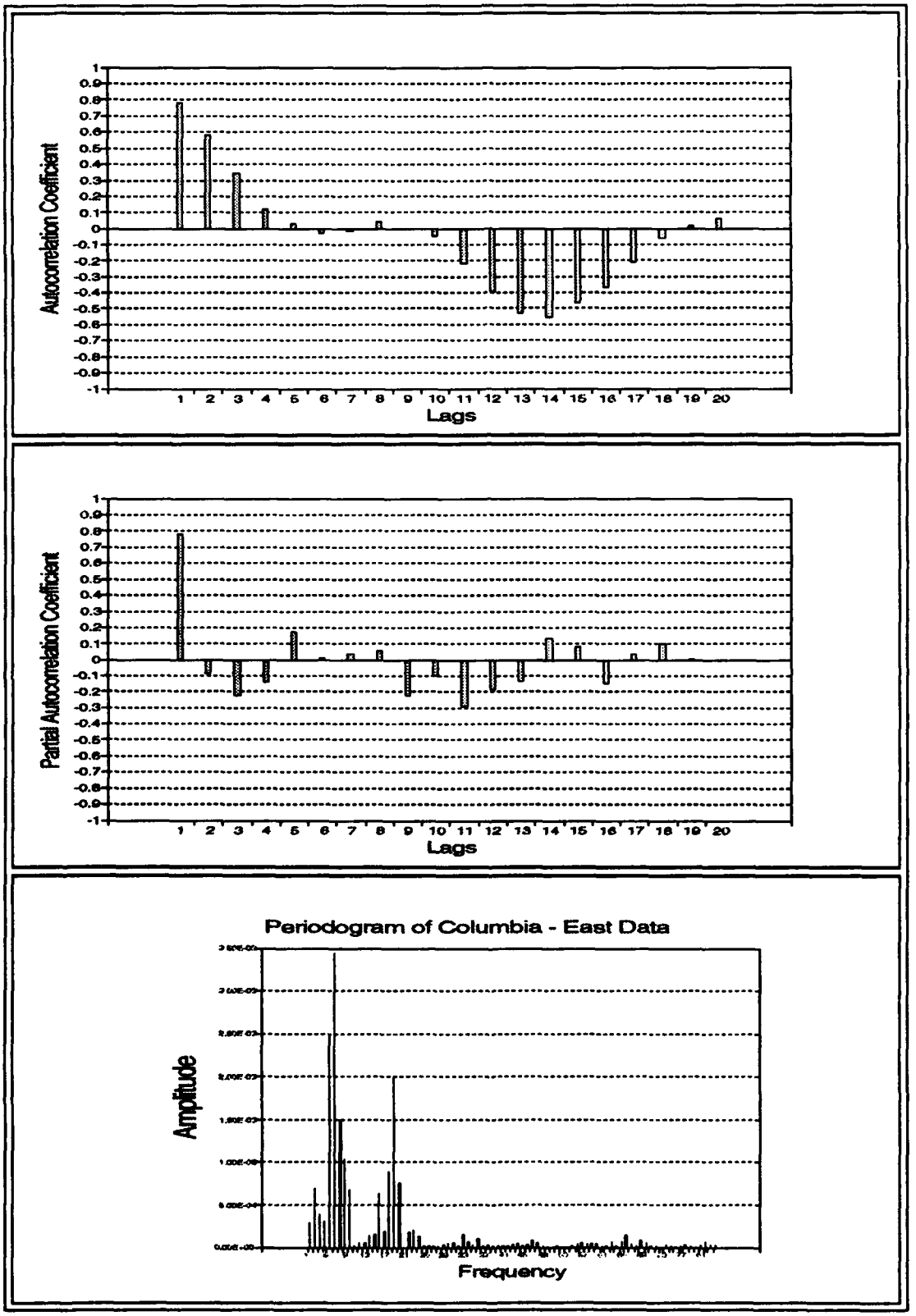

Figure F.10. ACF and PACF plots for the east azimuth at Columbia, Missouri.

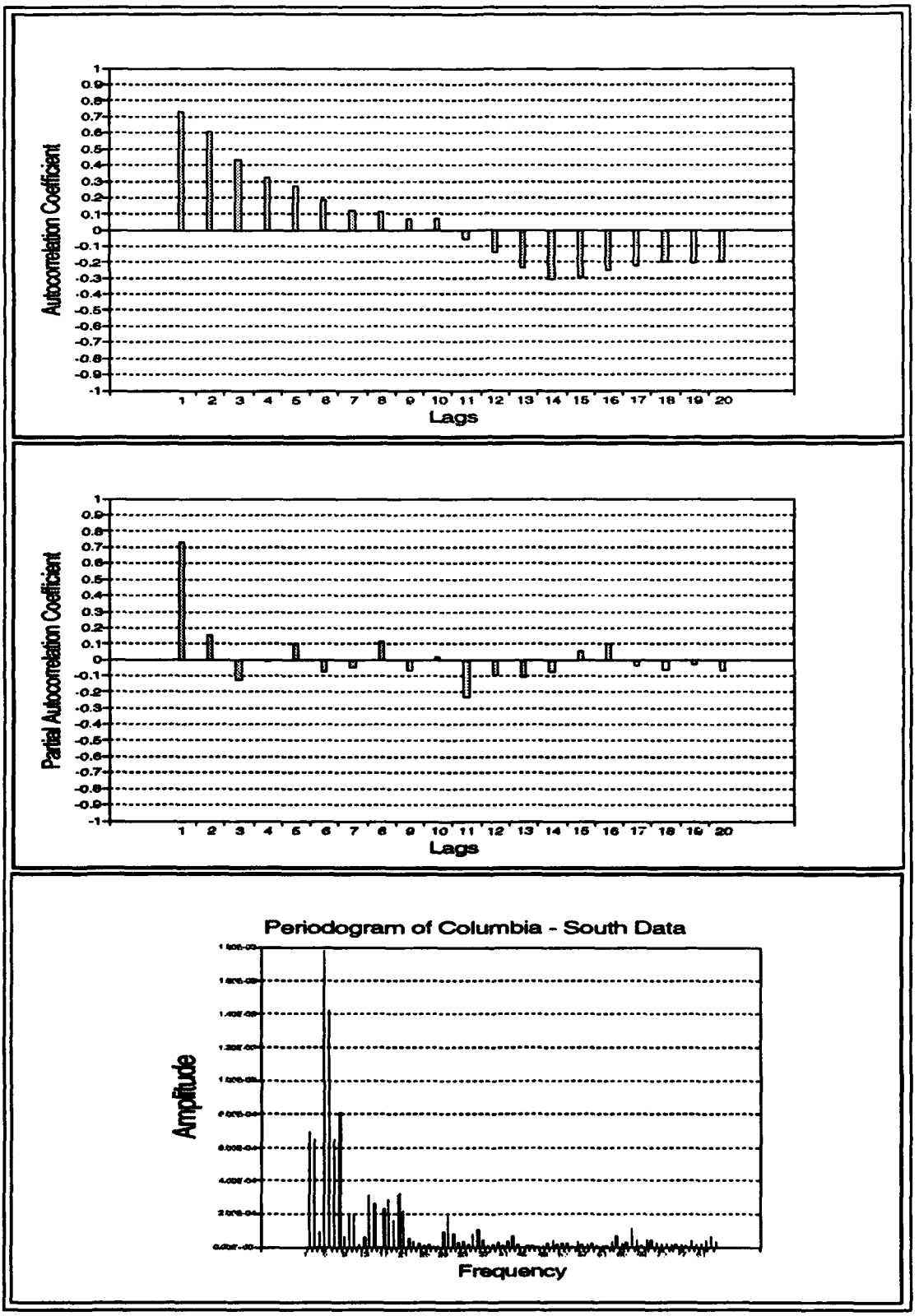

Figure F.11. ACF and PACF plots for the south azimuth at Columbia, Missouri.

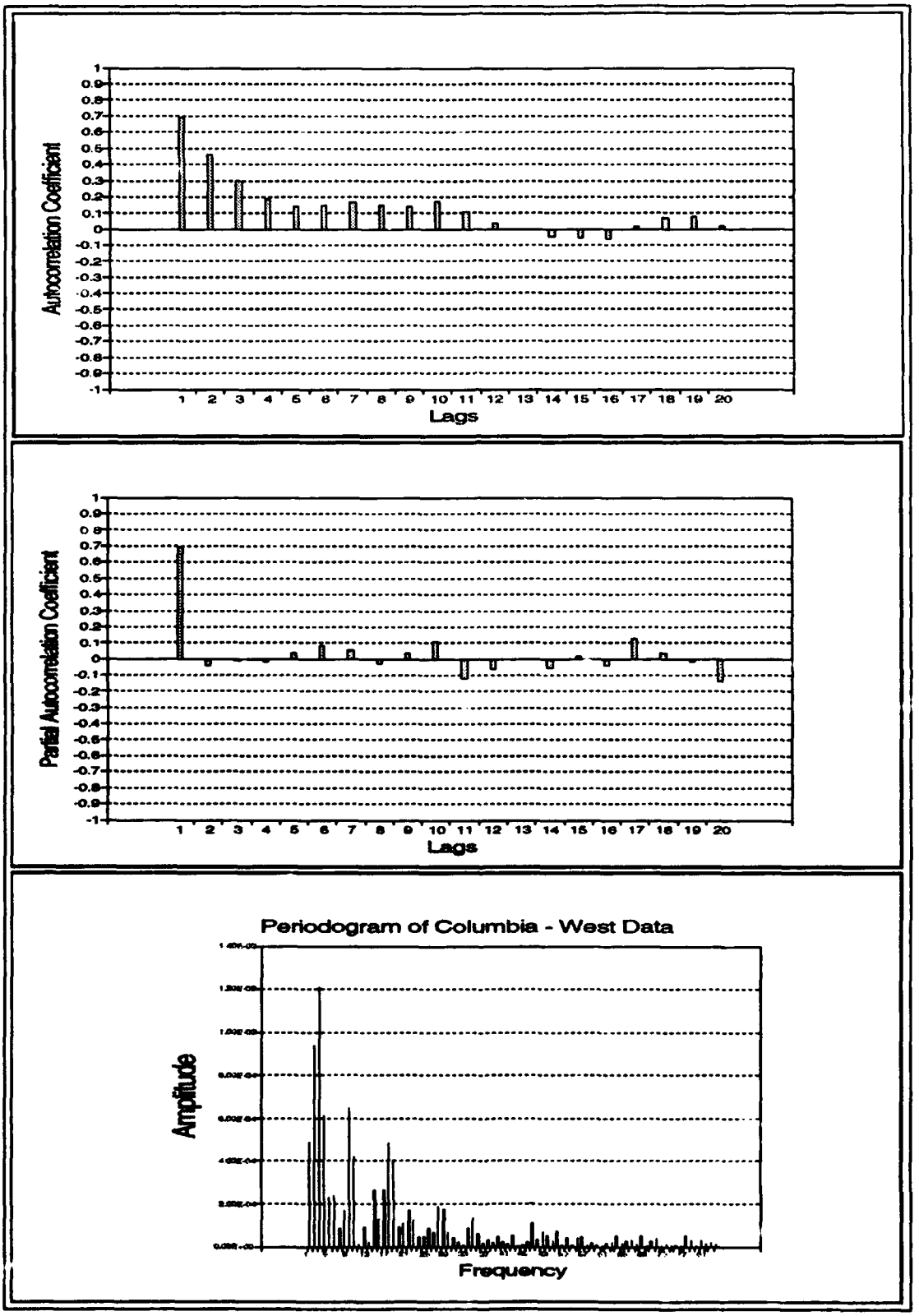

Figure F.12. ACF and PACF plots for the west azimuth at Columbia, Missouri.

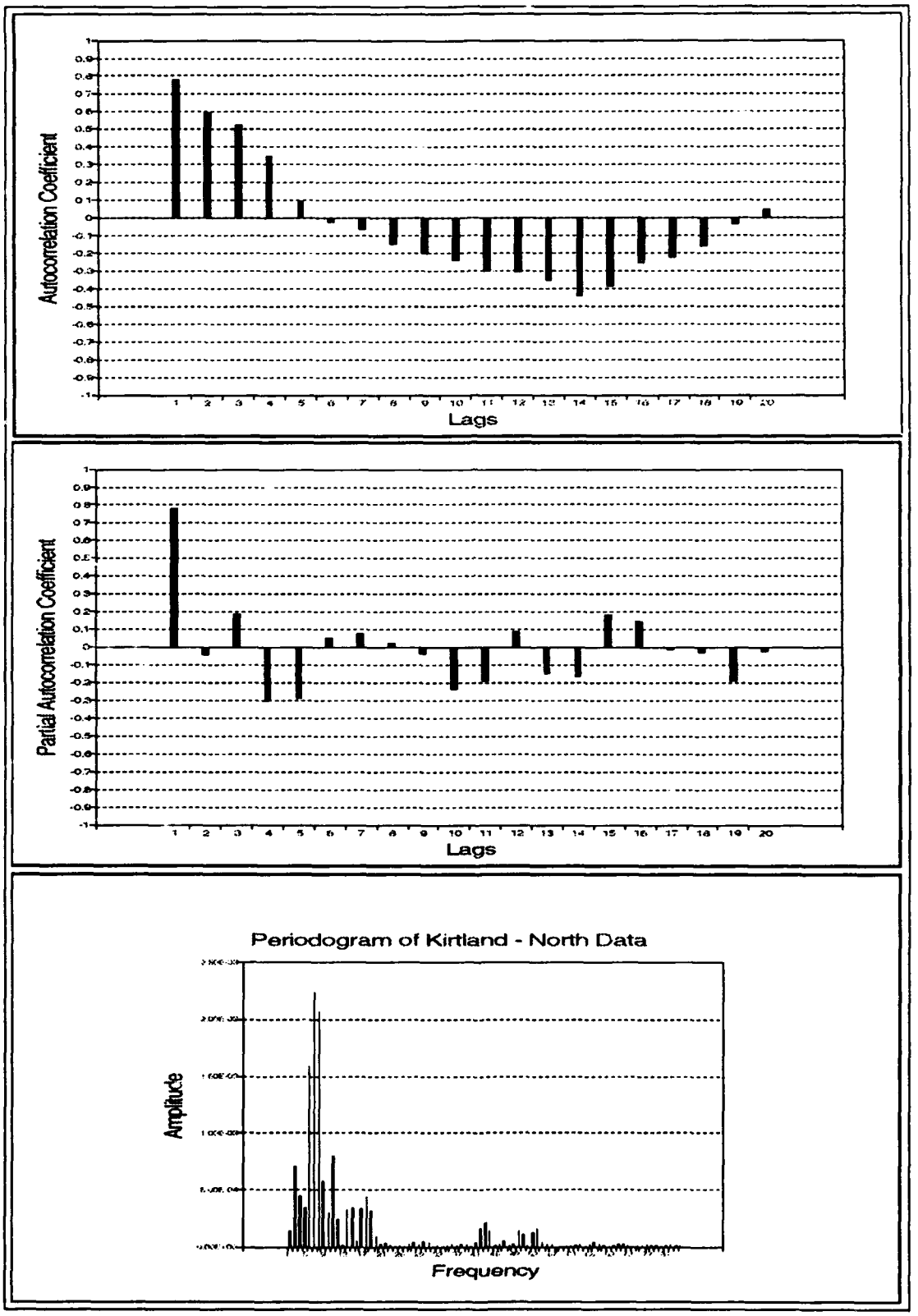

Figure F.13. ACF and PACF plots for the north azimuth at Kirtland AFB, New Mexico.

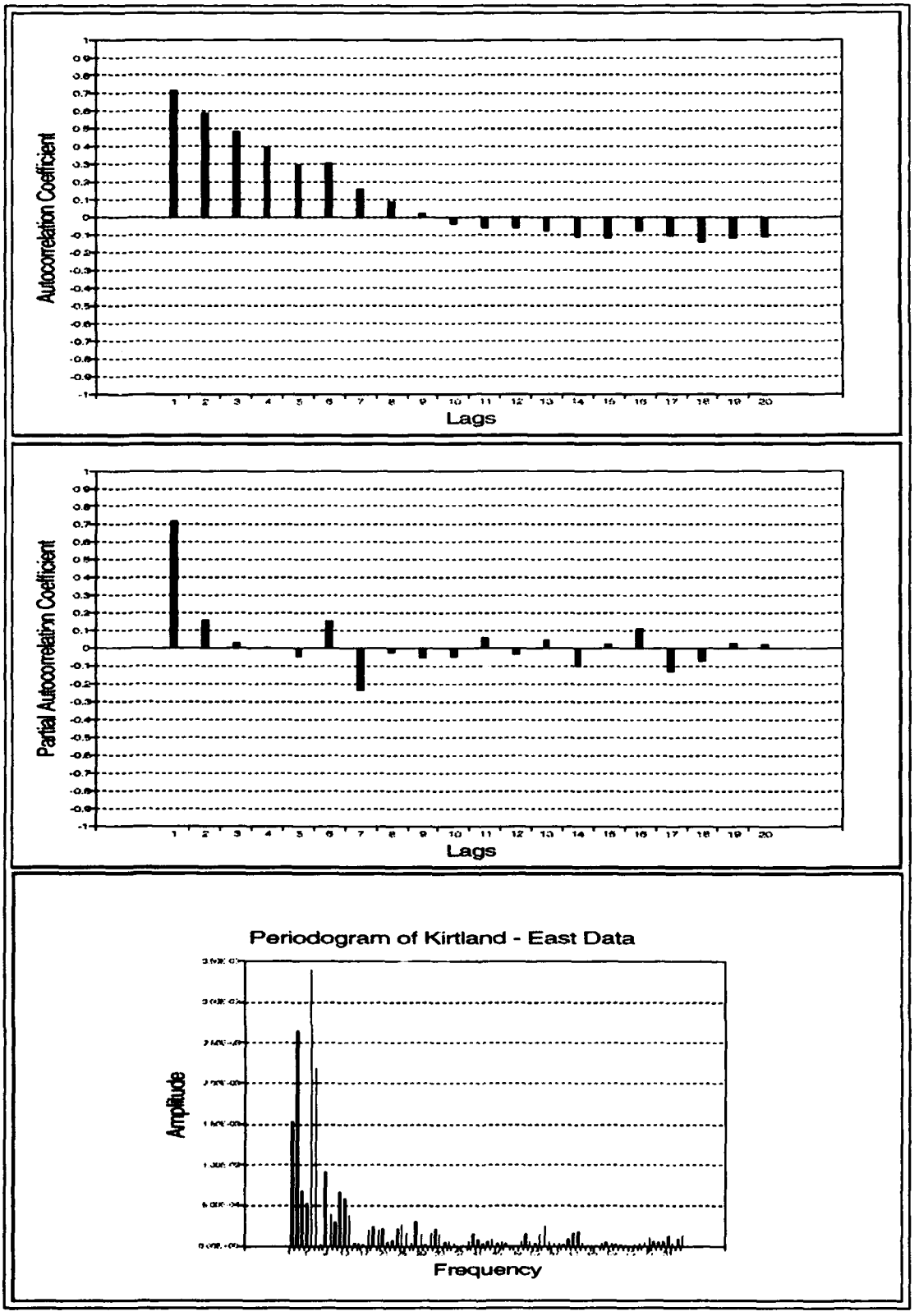

Figure F.14. ACF and PACF plots for the east azimuth at Kirtland AFB, New Mexico.

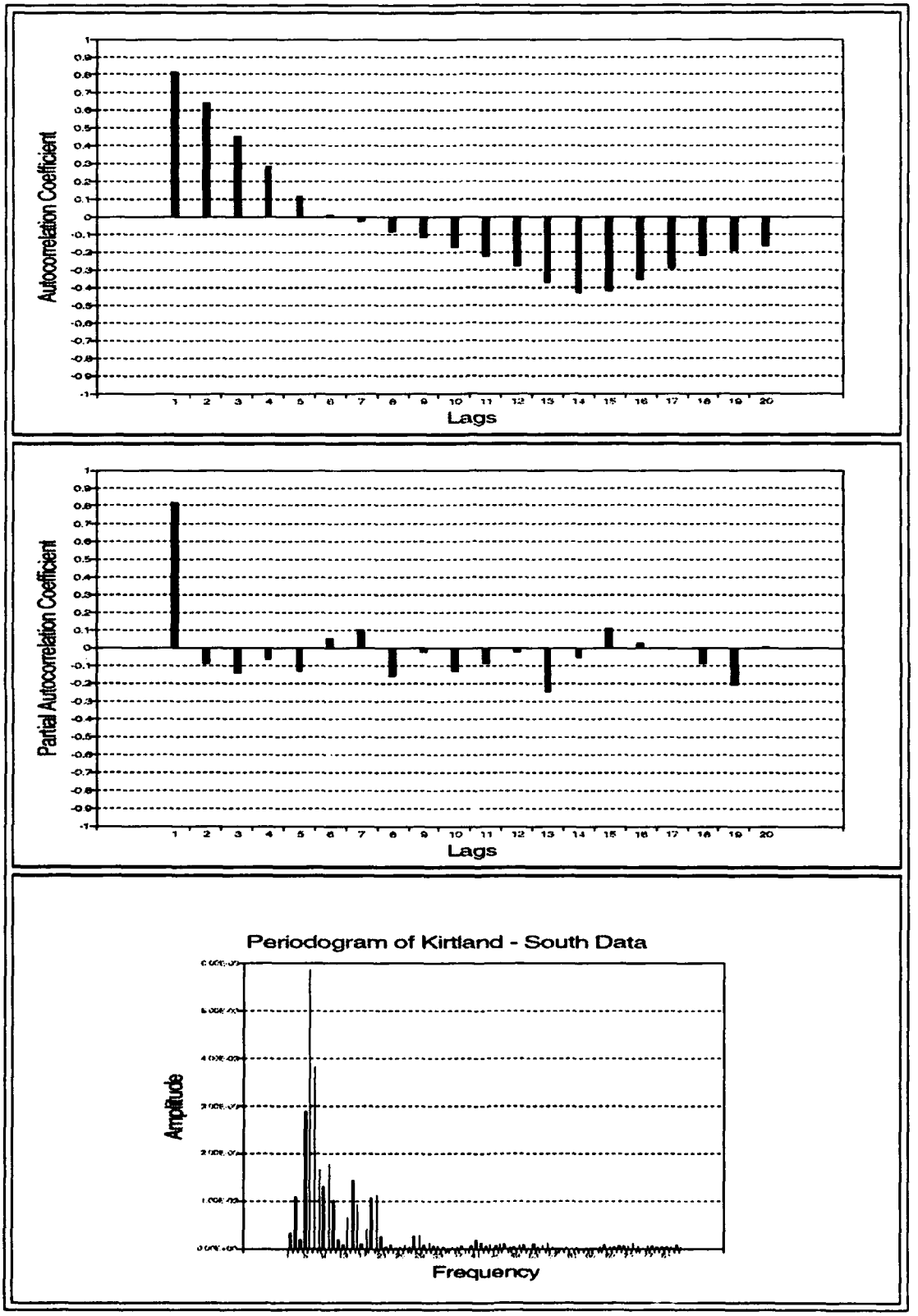

Figure F.15. ACF and PACF plots for the south azimuth at Kirtland AFB, New Mexico.

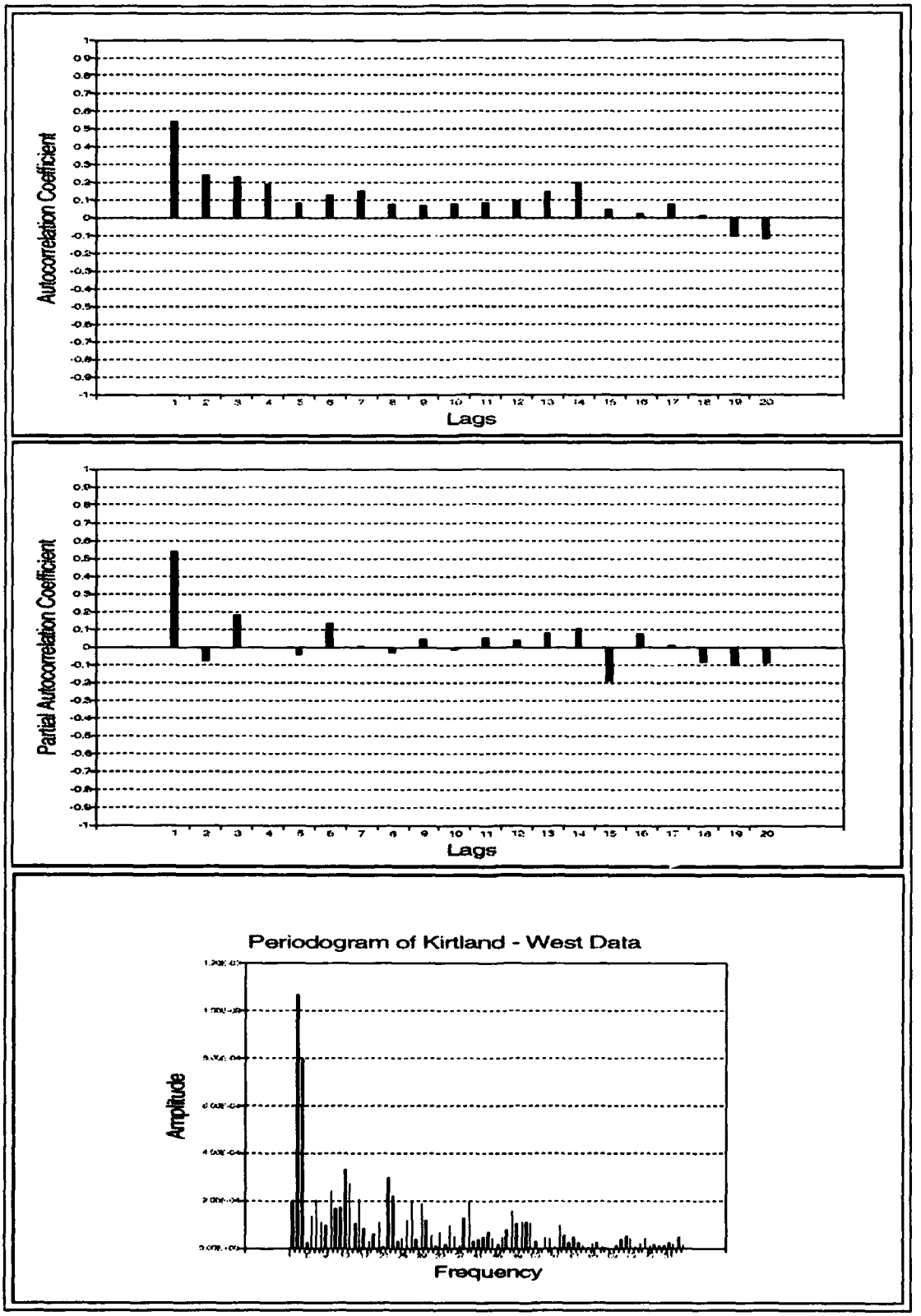

Figure F.16. ACF and PACF plots for the west azimuth at Kirtland AFB, New Mexico.

Plots of ACFs and PACFs for Residuals (by model)

The ACF and PACF for the residuals of each model are displayed in this section. The data is presented in order according to the site, direction, and model. Columbia data is presented first and is followed by the Kirtland data. The order of directions is north, east, south, west. The models follow the sequence (0,0,1), (0,0,2), (1,0,0), (1,0,1), (1,0,2), (2,0,0), (2,0,1), and (2,0,2). For example, the first data presented is the Columbia North ACF and PACF for model (0,0,1) and is abbreviated CN(001).

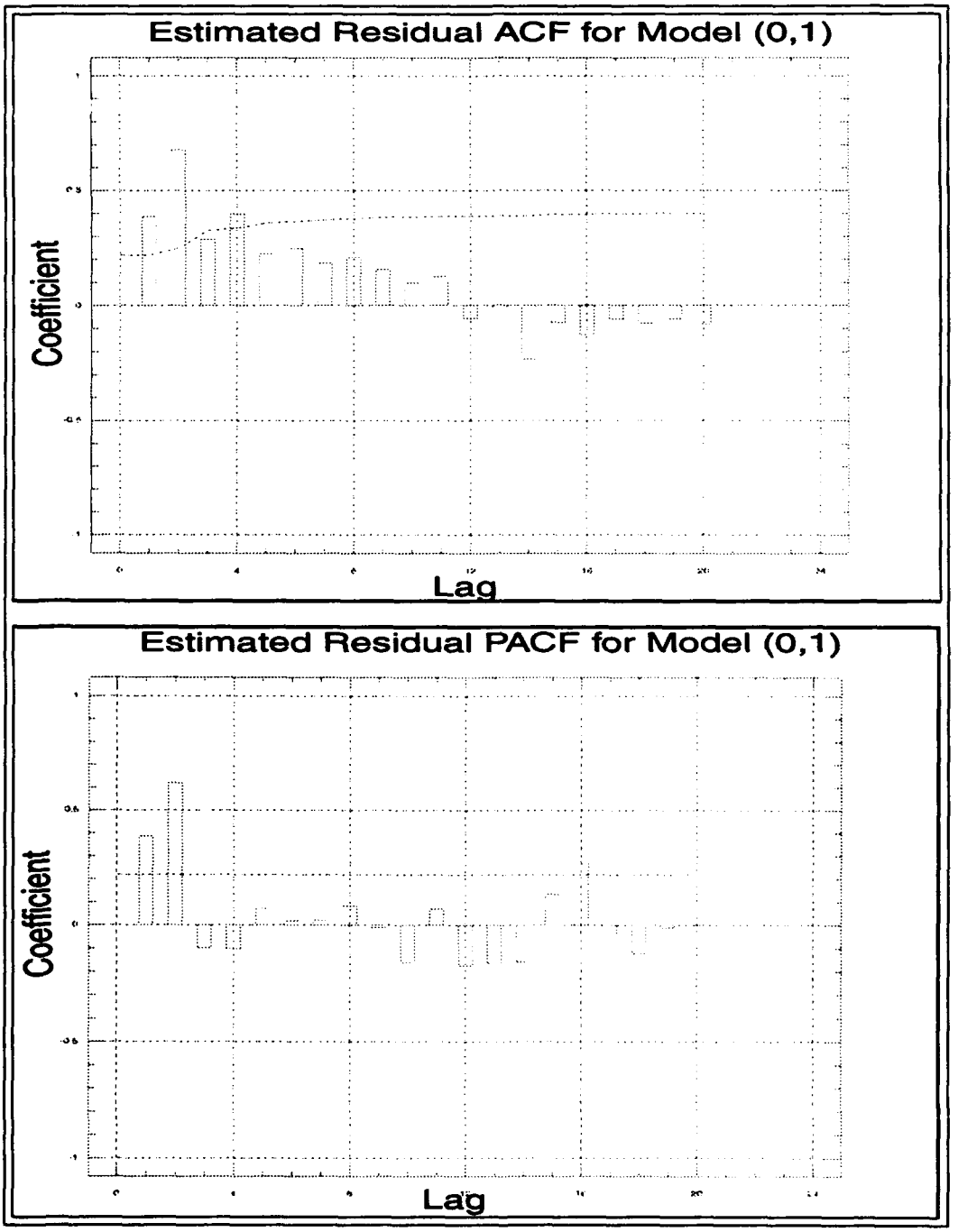

Figure F.17. Plots of the residual ACF and PACF for Model CN(0,0,1).

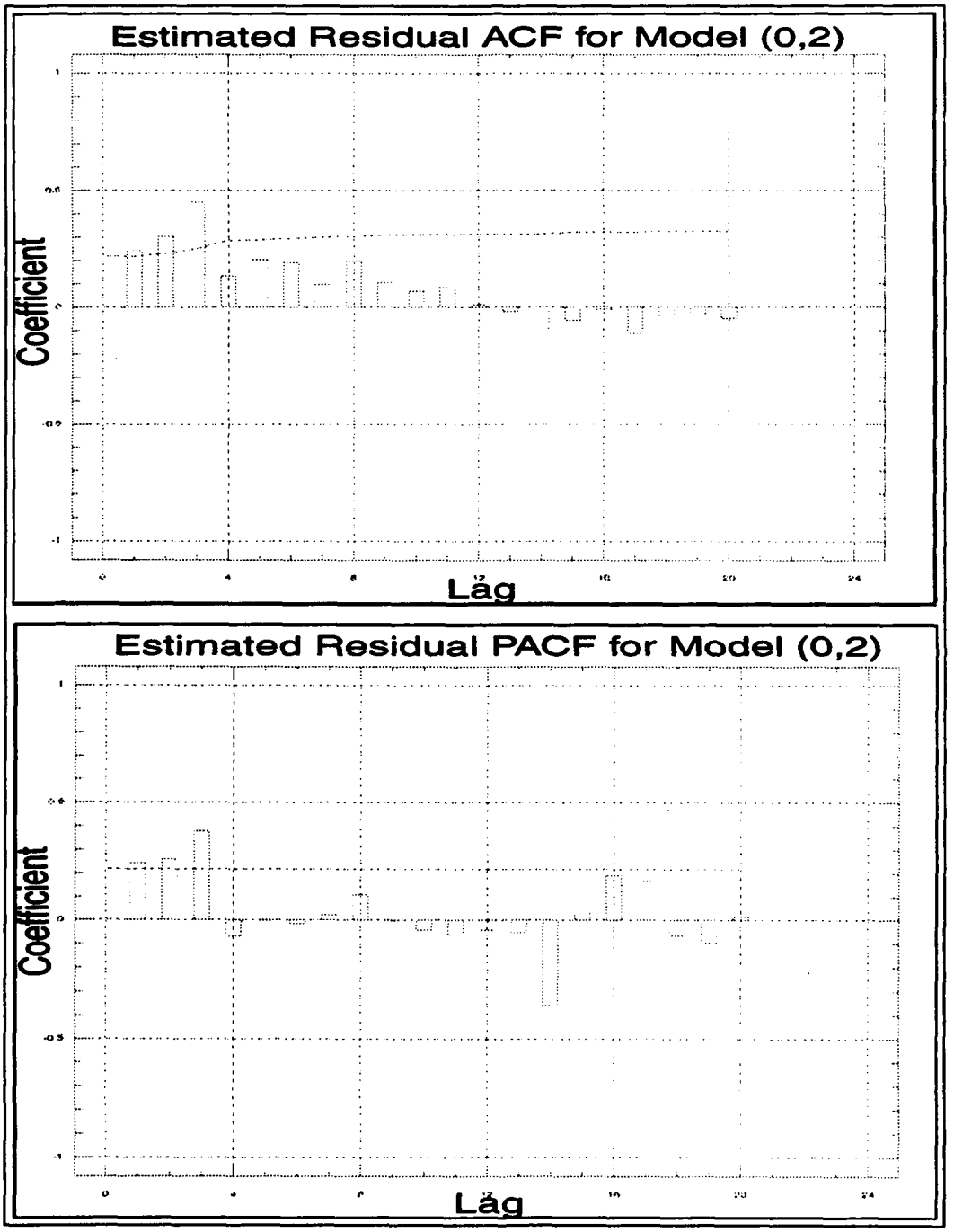

Figure F.18. Plots of the residual ACF and PACF for Model CN(0,0,2).

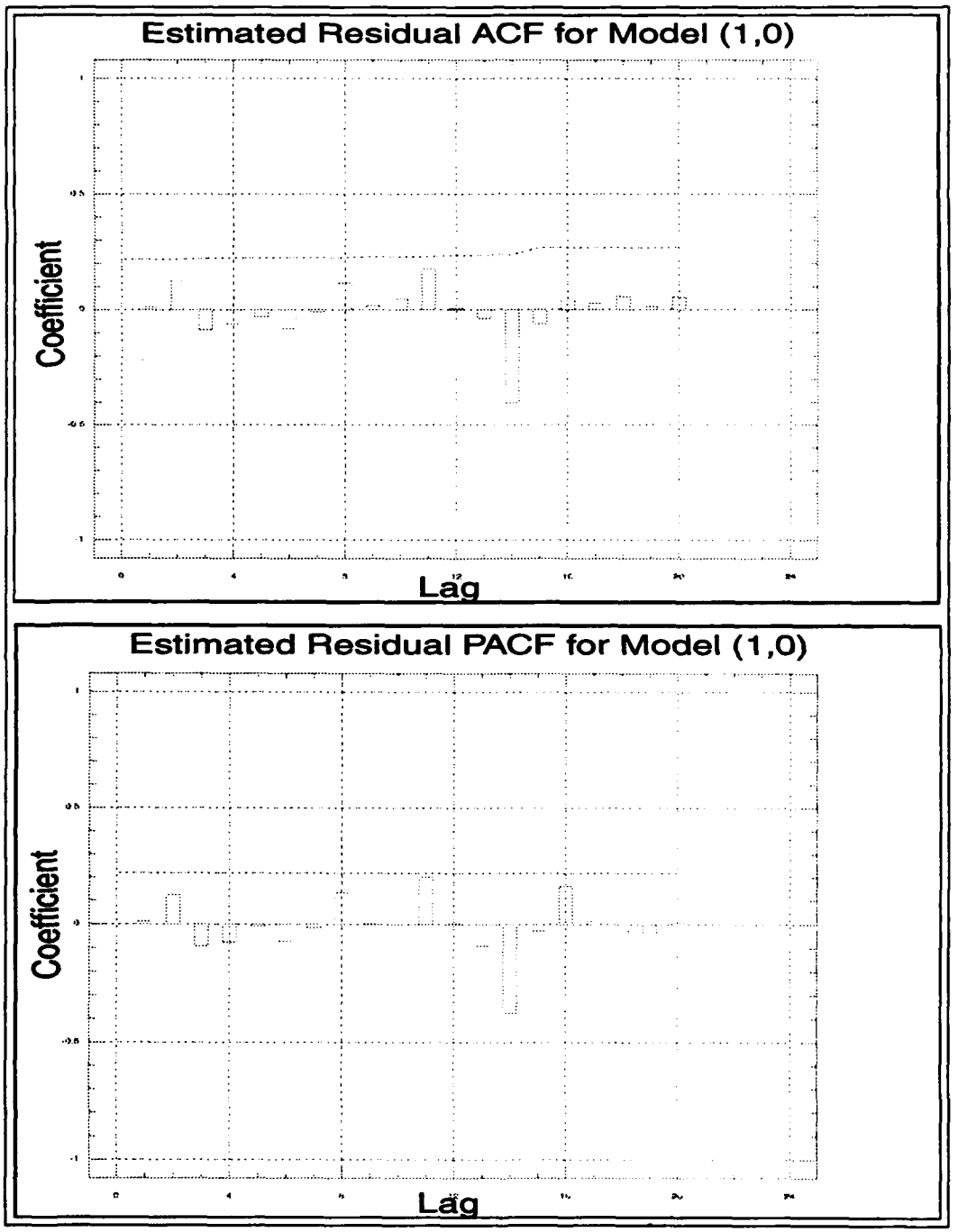

Figure F.19. Plots of the residual ACF and PACF for Model CN(1,0,0).

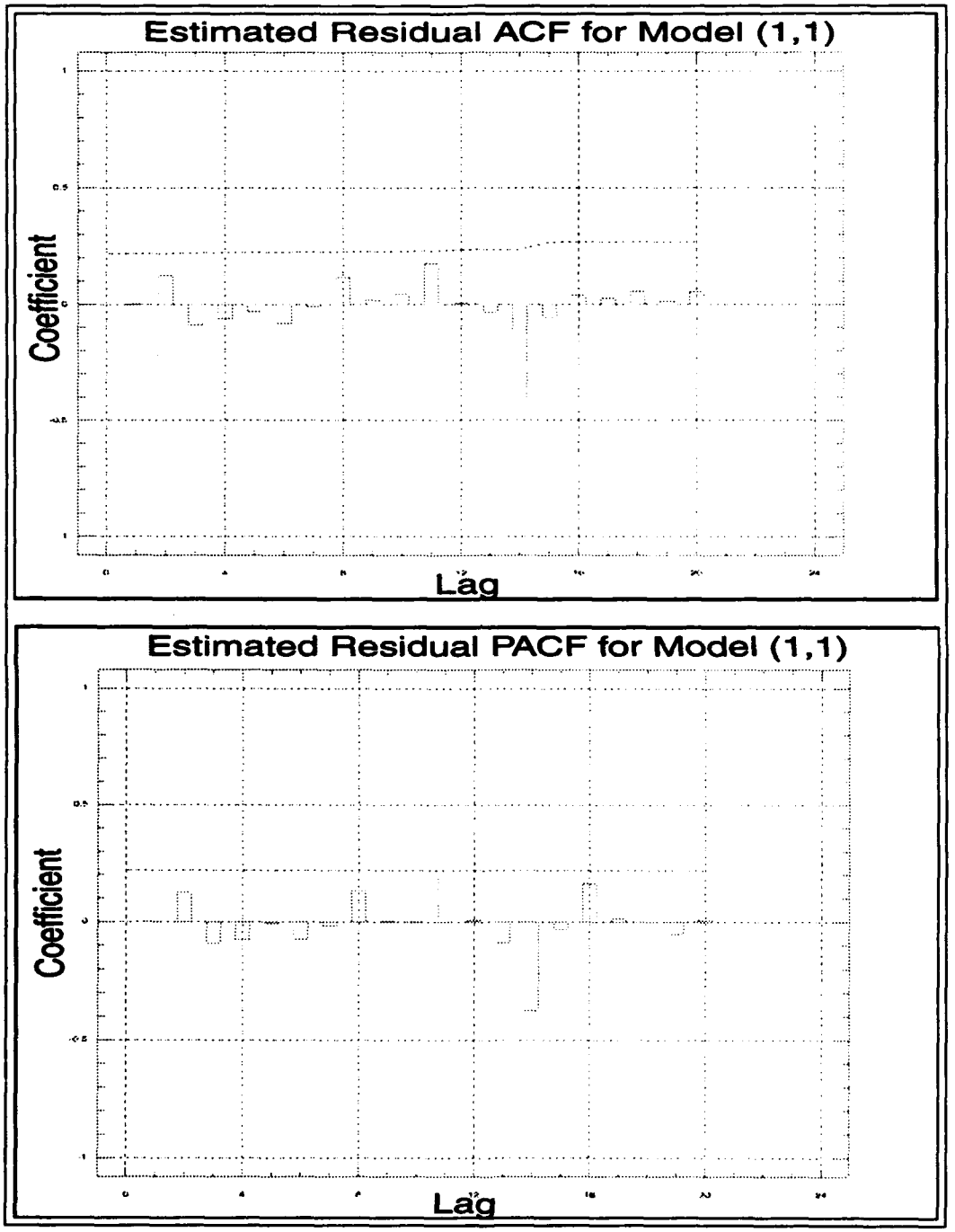

Figure F.20. Plots of the residual ACF and PACF for Model CN(1,0,1).

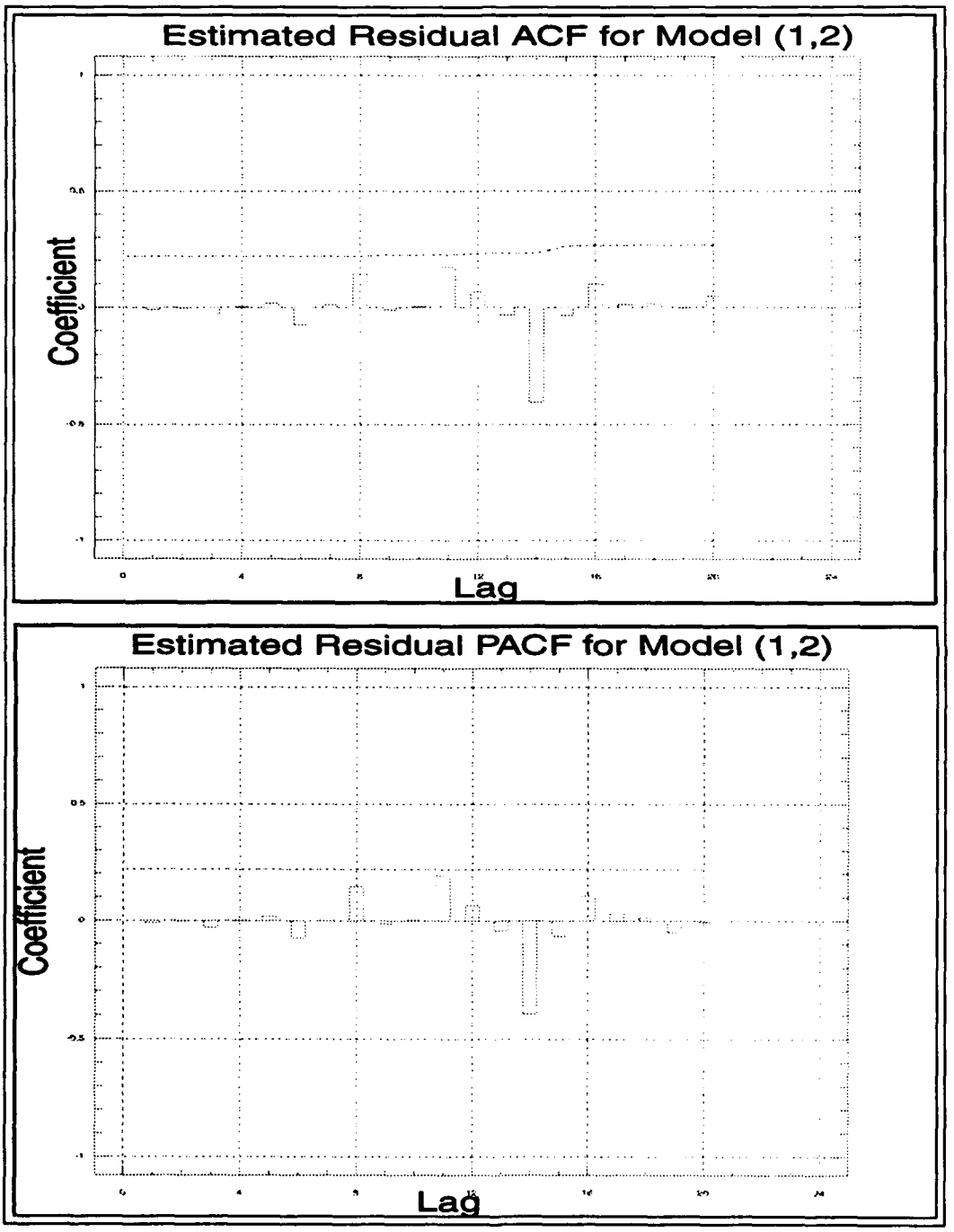

Figure F.21. Plots of the residual ACF and PACF for Model CN(1,0,2).

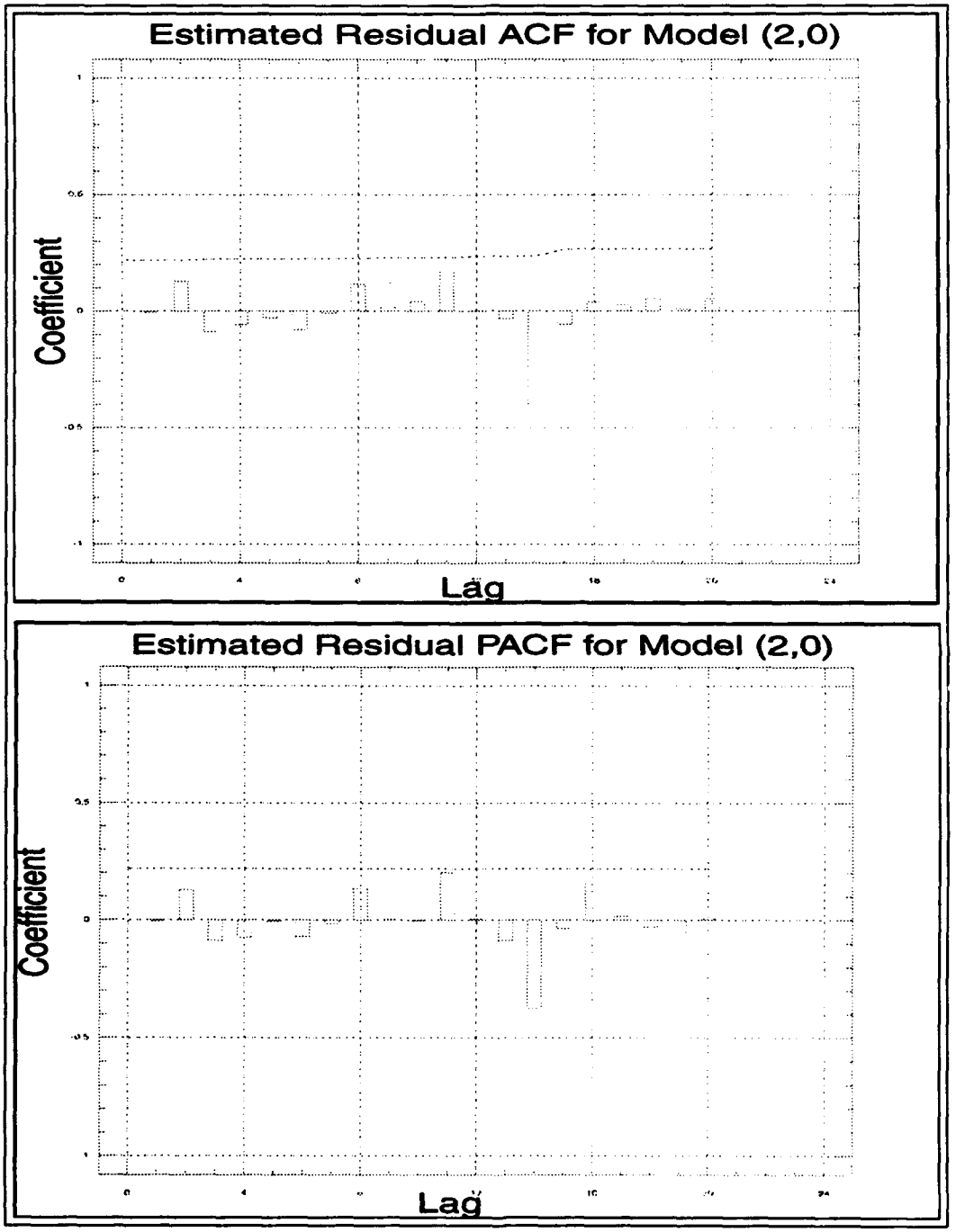

Figure F.22. Plots of the residual ACF and PACF for Model CN(2,0,0).

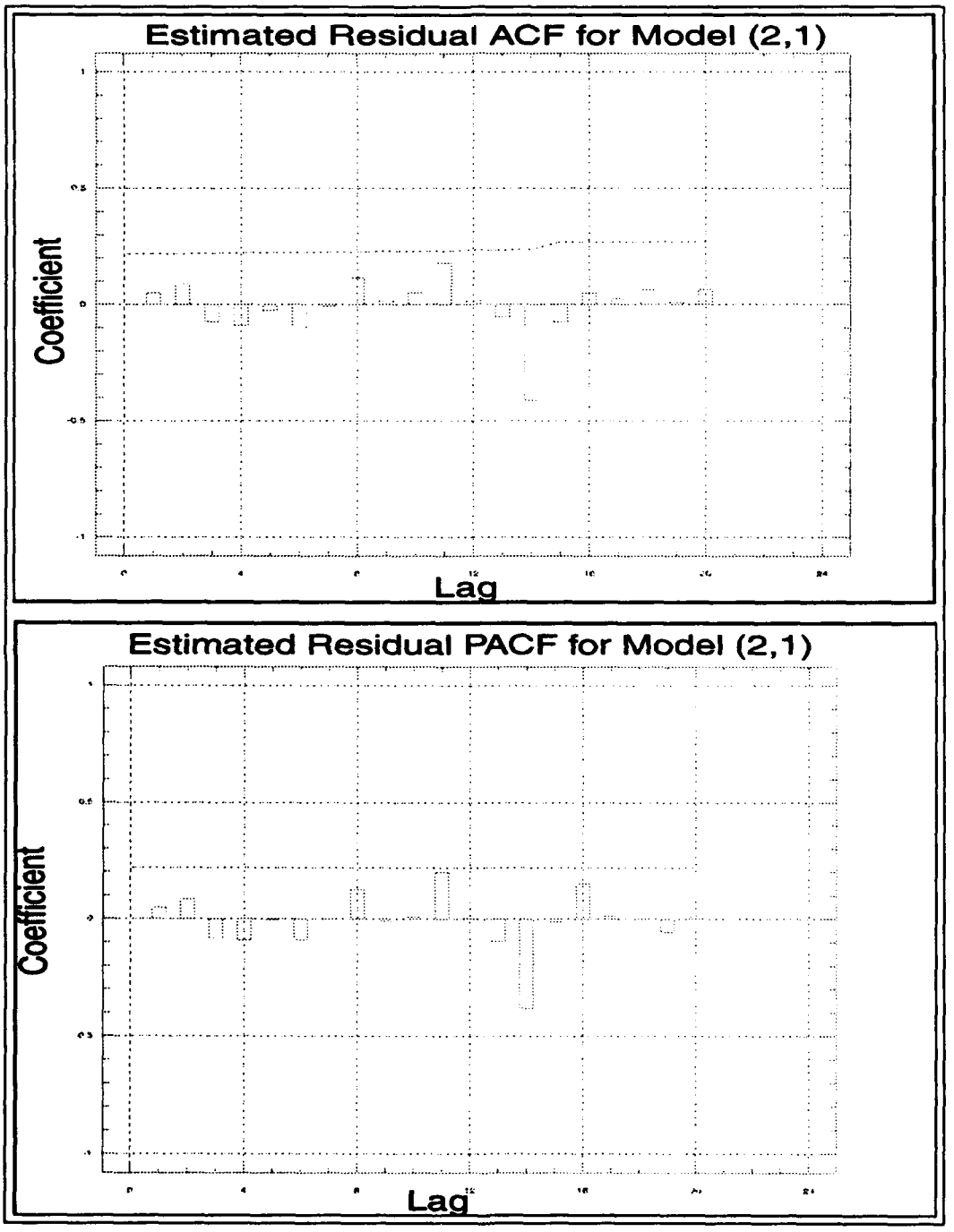

Figure F.23. Plots of the residual ACF and PACF for Model CN(2,0,1).

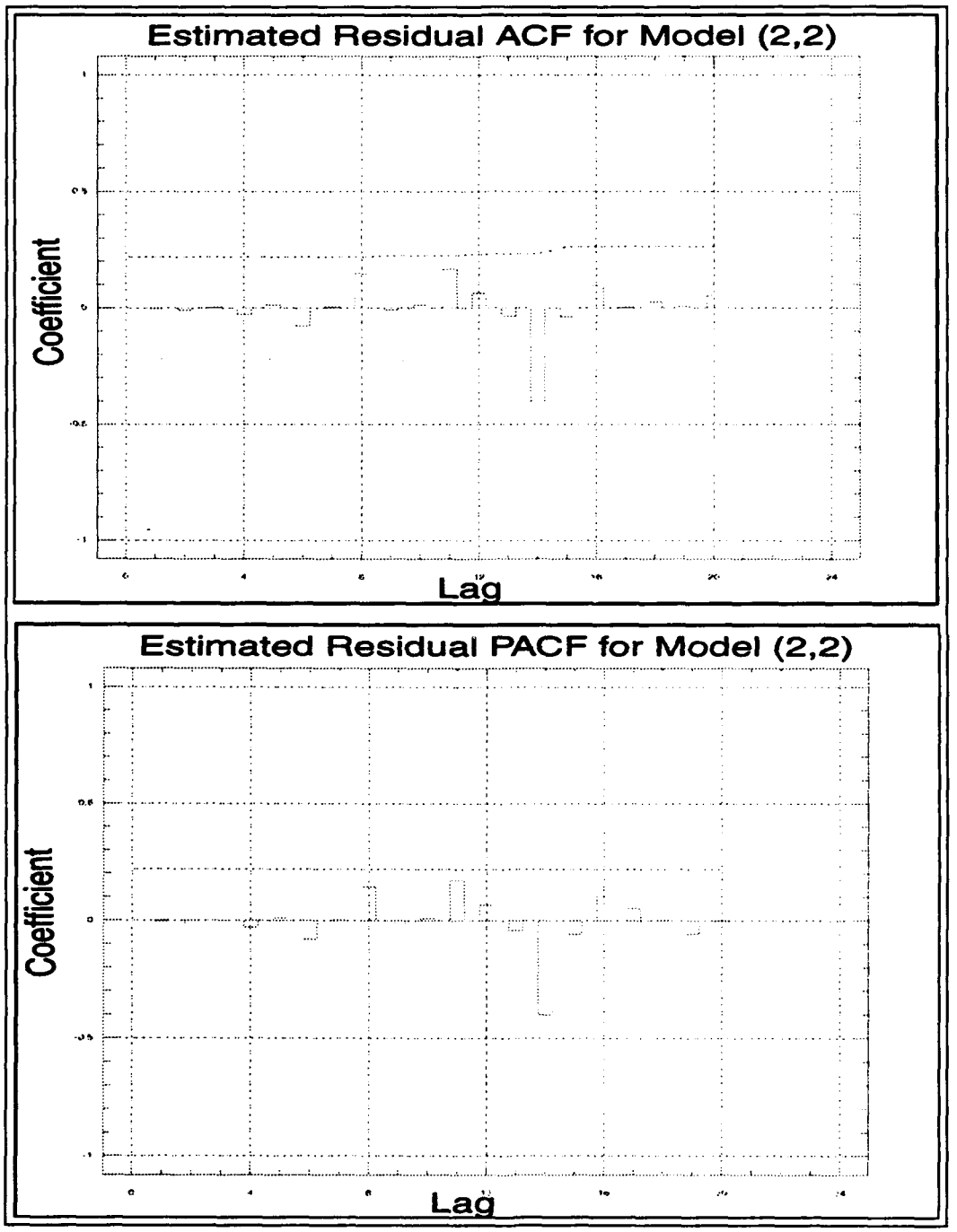

Figure F.24. Plots of the residual ACF and PACF for Model CN(2,0,2).

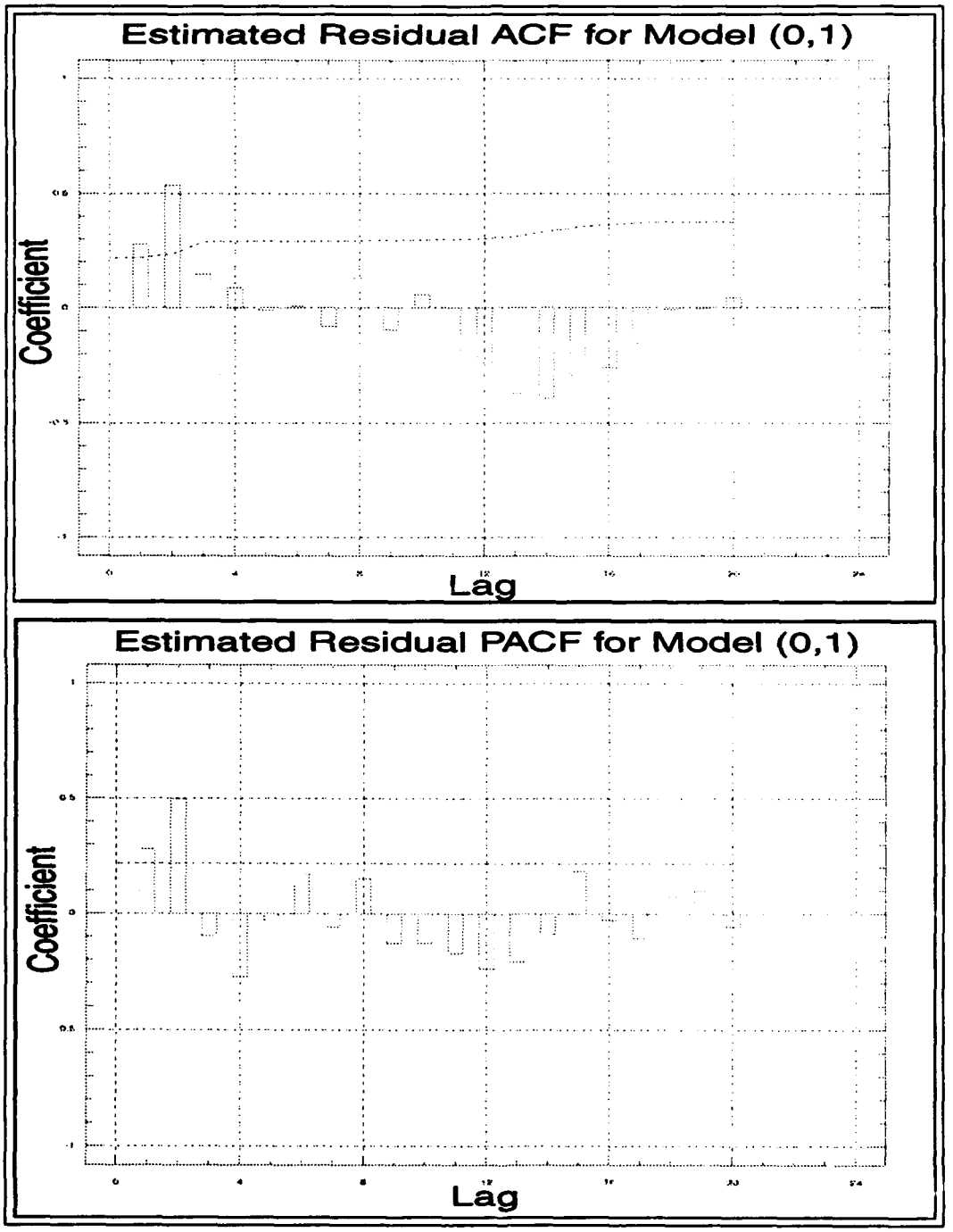

Figure F.25. Plots of the residual ACF and PACF for Model CE(0,0,1).

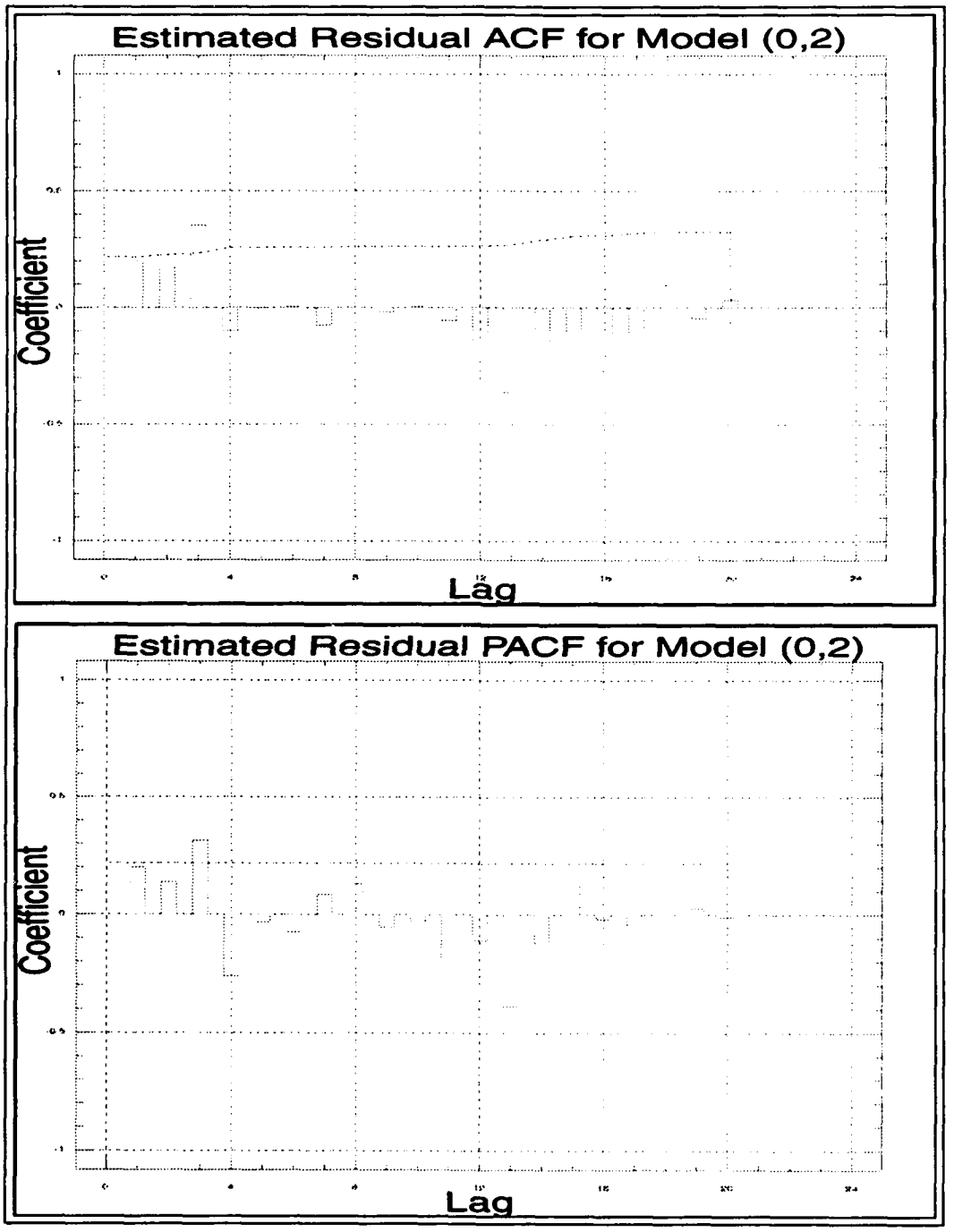

Figure F.26. Plots of the residual ACF and PACF for Model CE(0,0,2).

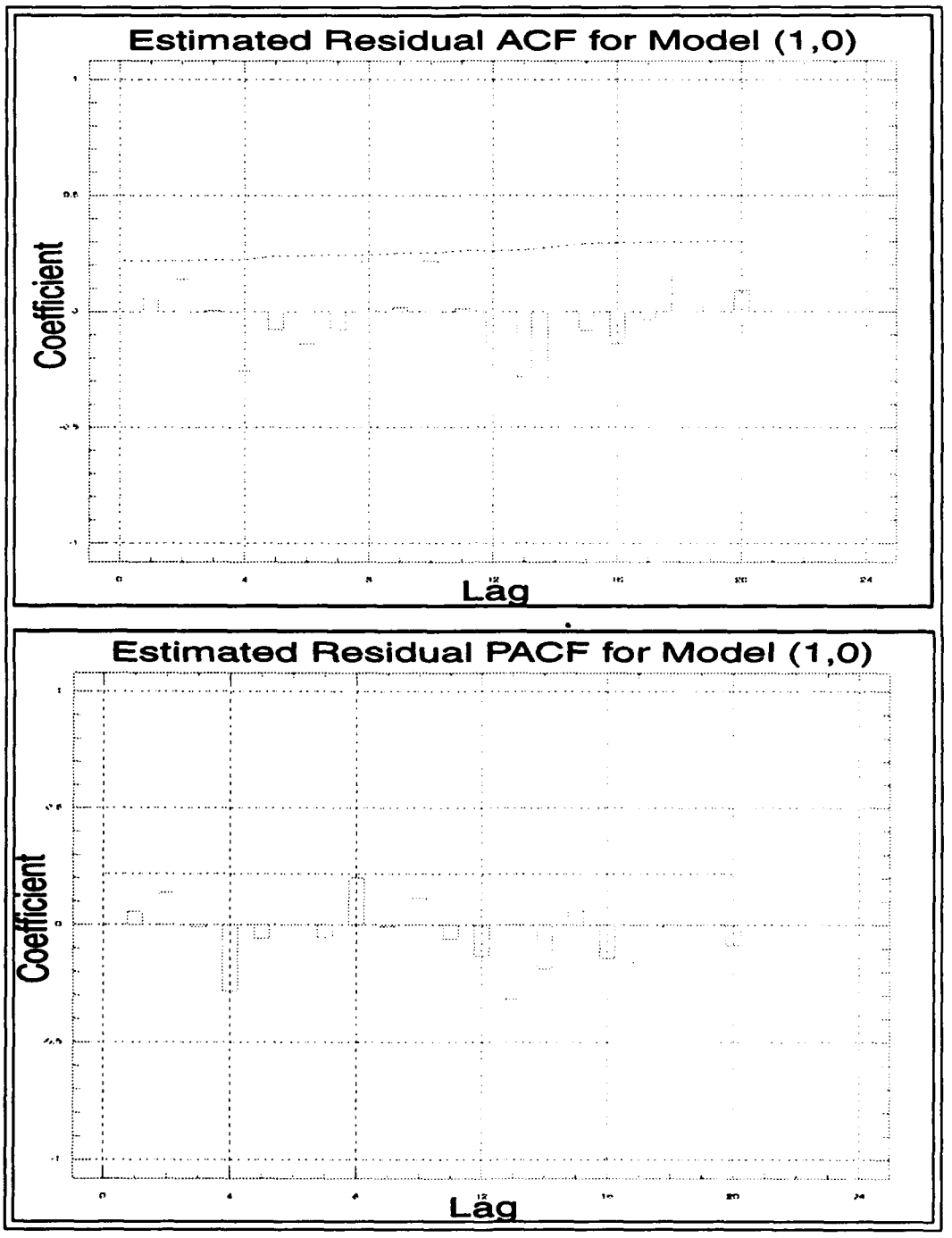

Figure F.27. Plots of the residual ACF and PACF for Model CE(1,0,0).

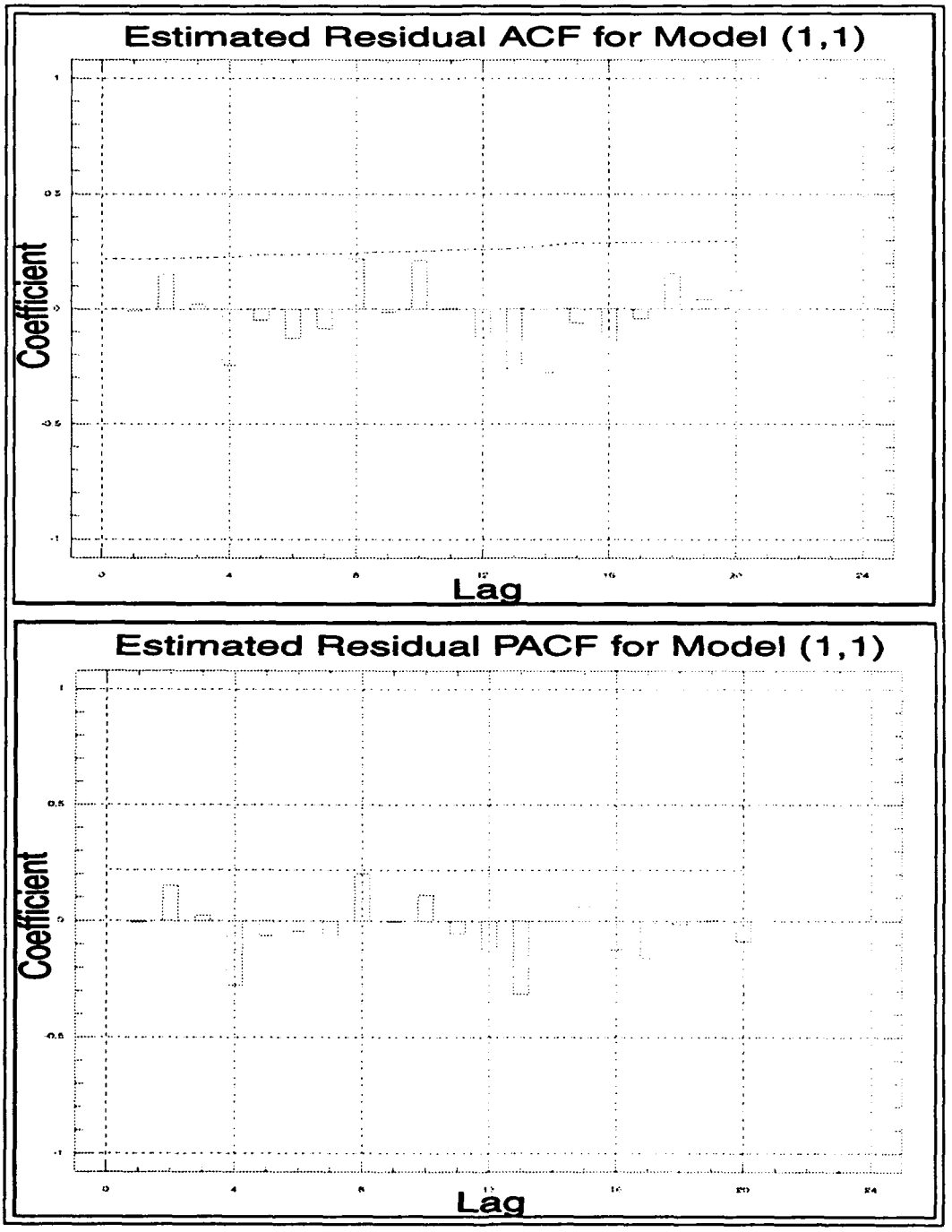

Figure F.28. Plots of the residual ACF and PACF for Model CE(1,0,1).

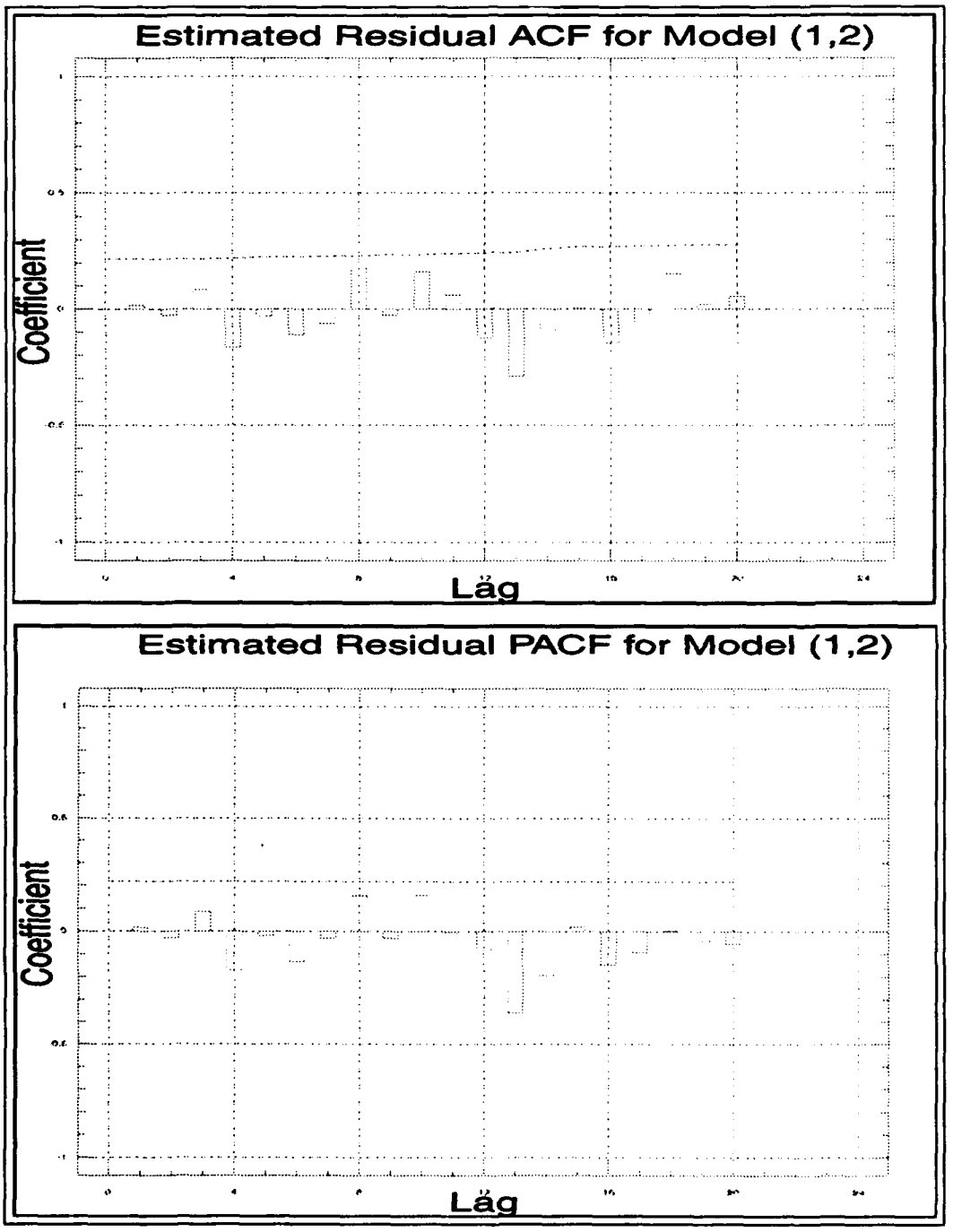

Figure F.29. Plots of the residual ACF and PACF for Model CE(1,0,2).

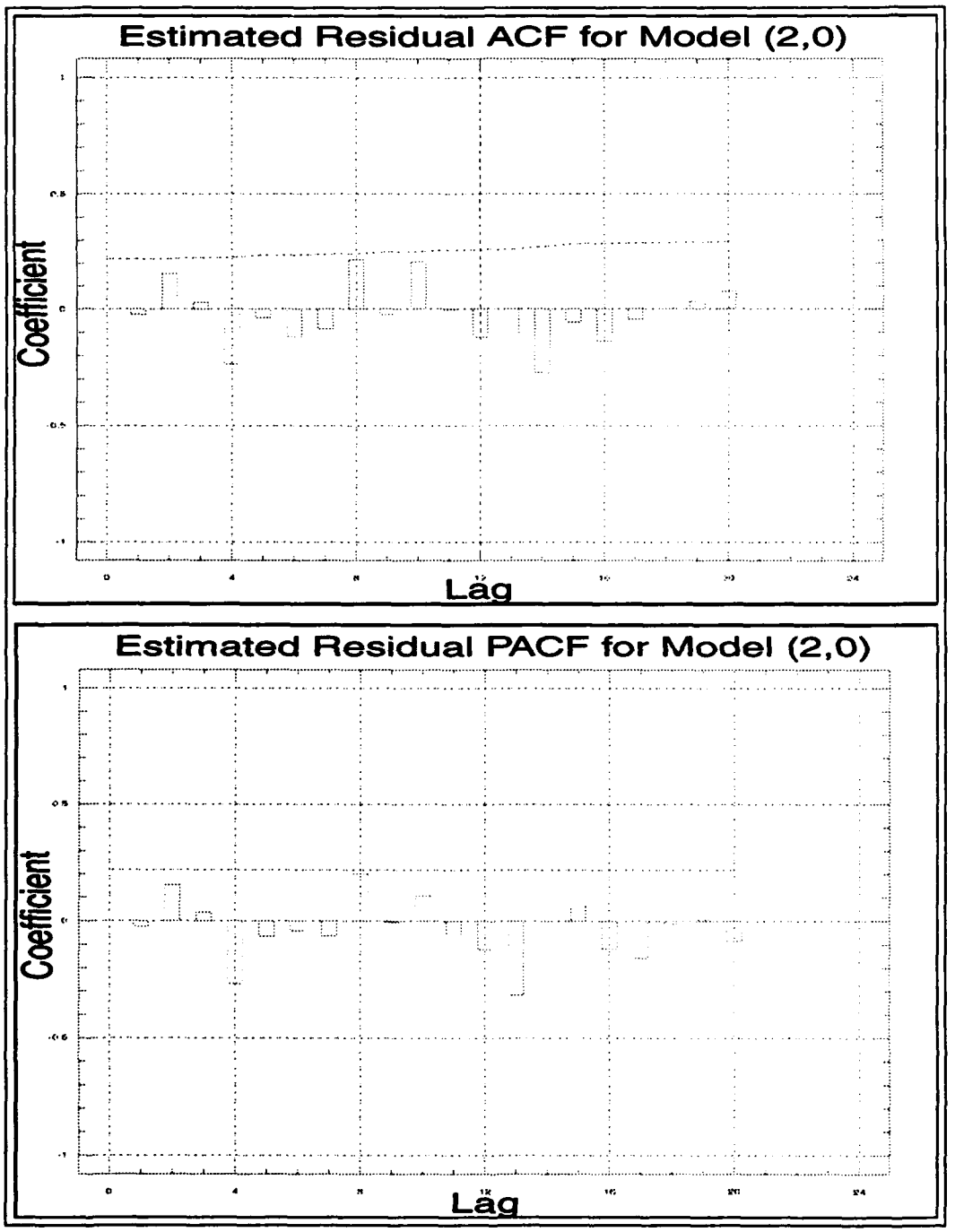

Figure F.30. Plots of the residual ACF and PACF for Model CE(2,0,0).

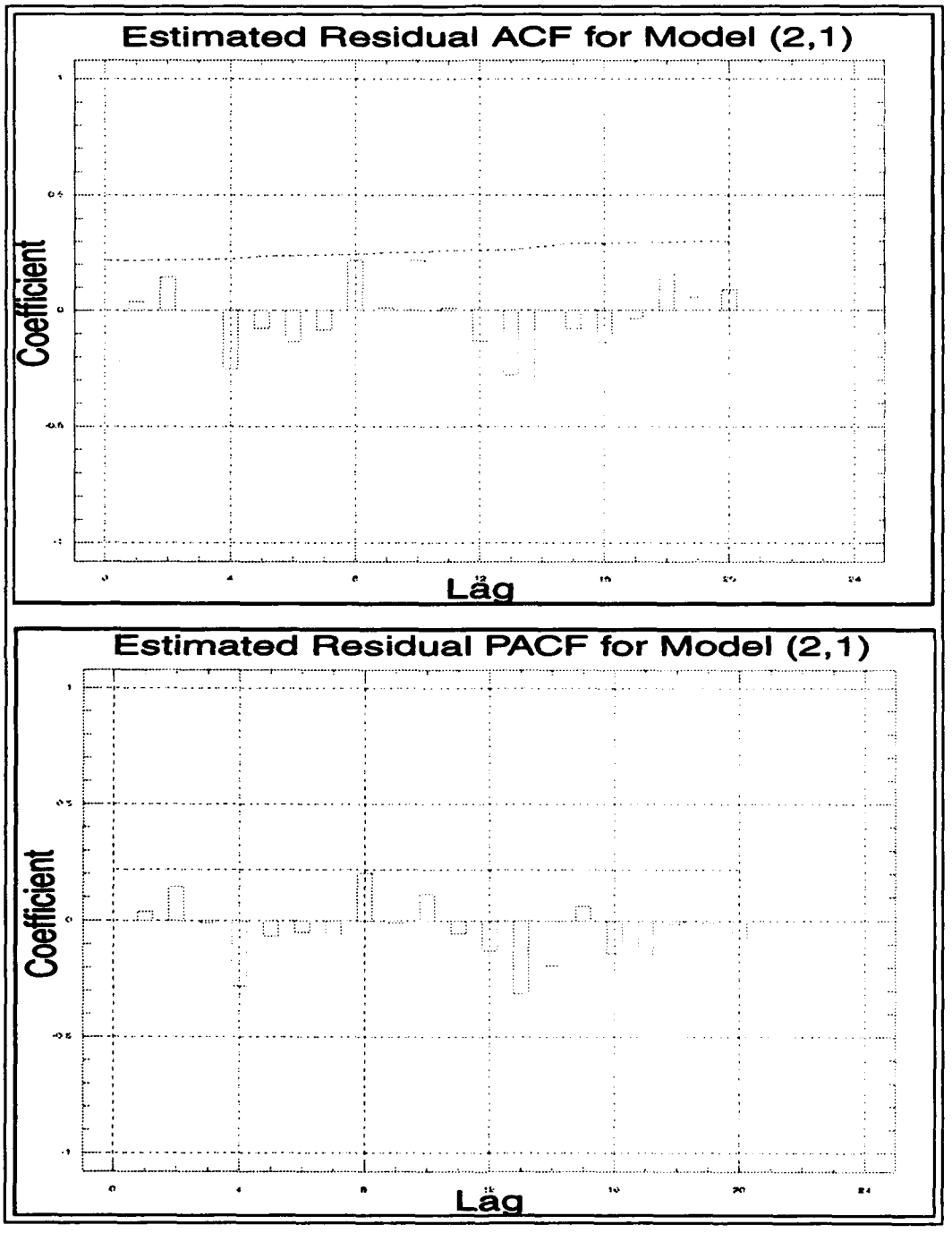

Figure F.31. Plots of the residual ACF and PACF for Model CE(2,0,1).

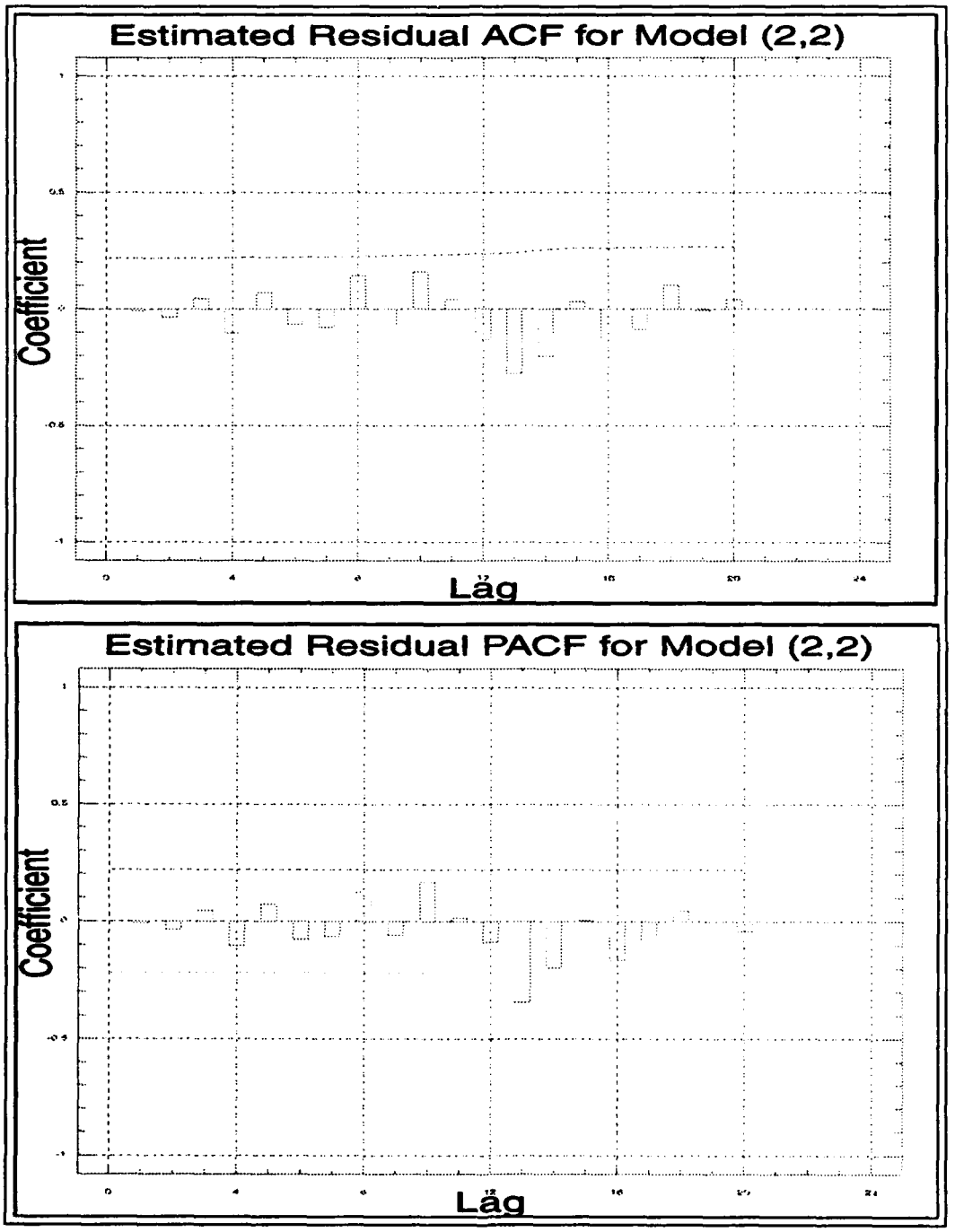

Figure F.32. Plots of the residual ACF and PACF for Model CE(2,0,2).

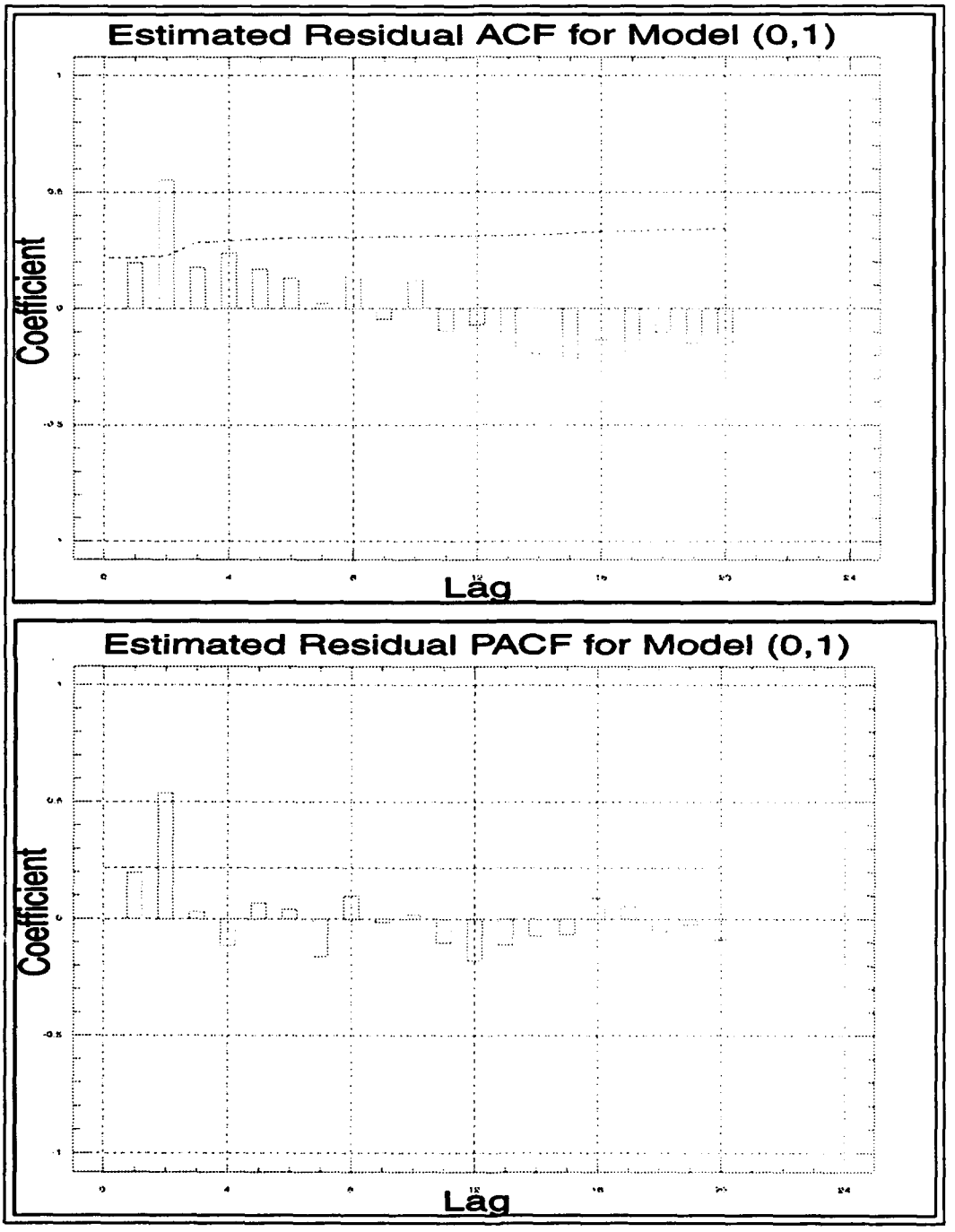

Figure F.33. Plots of the residual ACF and PACF for Model CS(0,0,1).

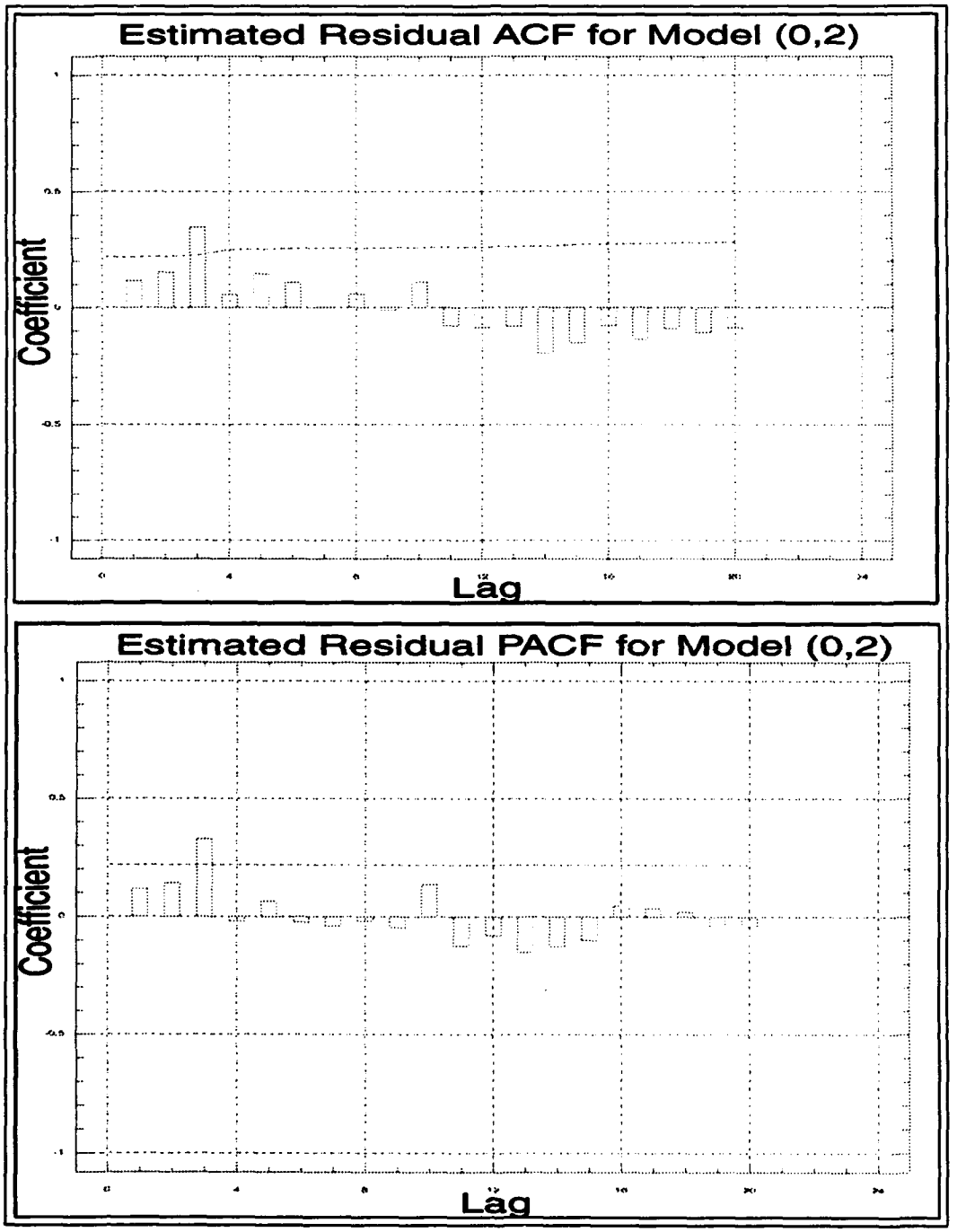

Figure F.34. Plots of the residual ACF and PACF for Model CS(0,0,2).

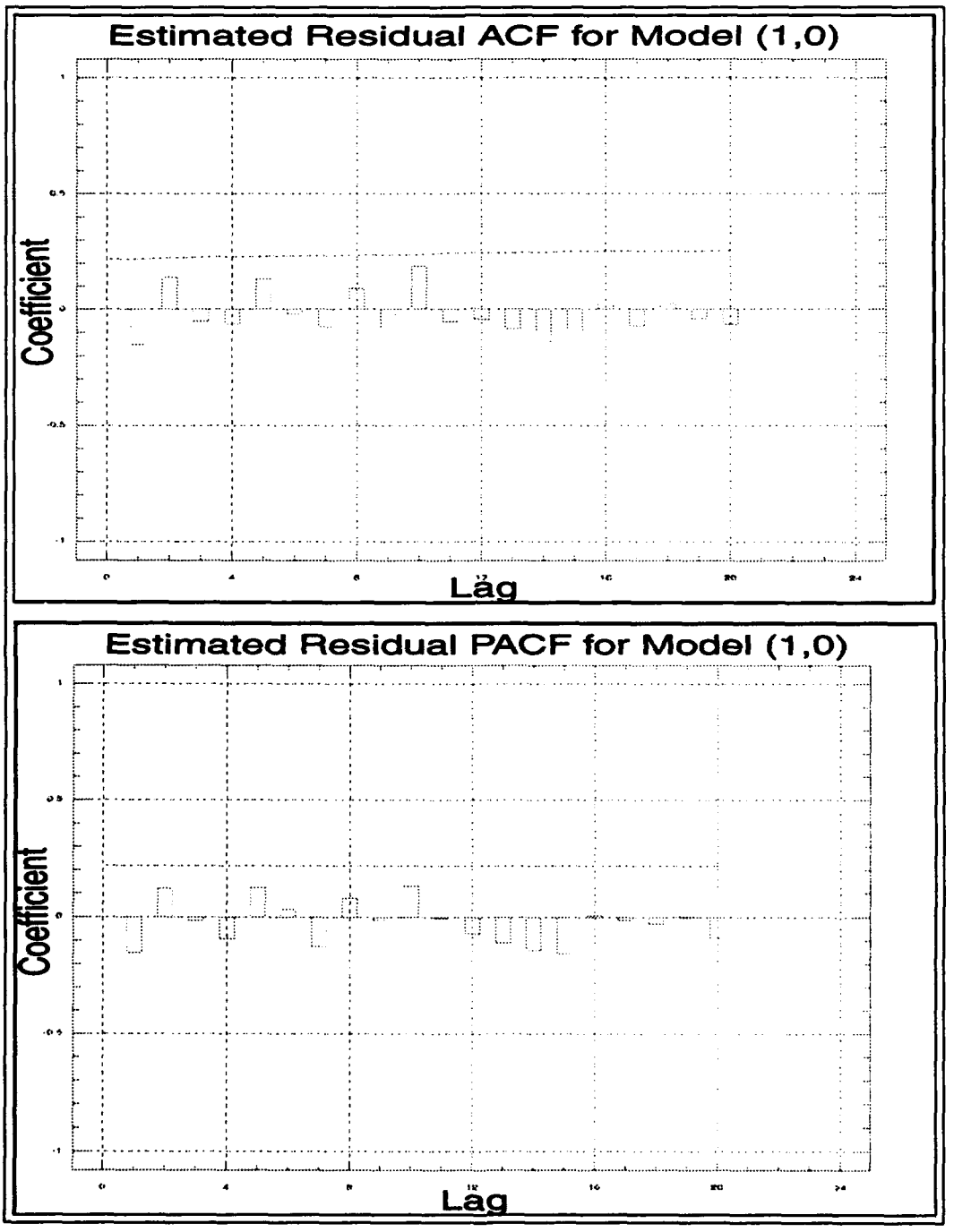

Figure F.35. Plots of the residual ACF and PACF for Model CS(1,0,0).

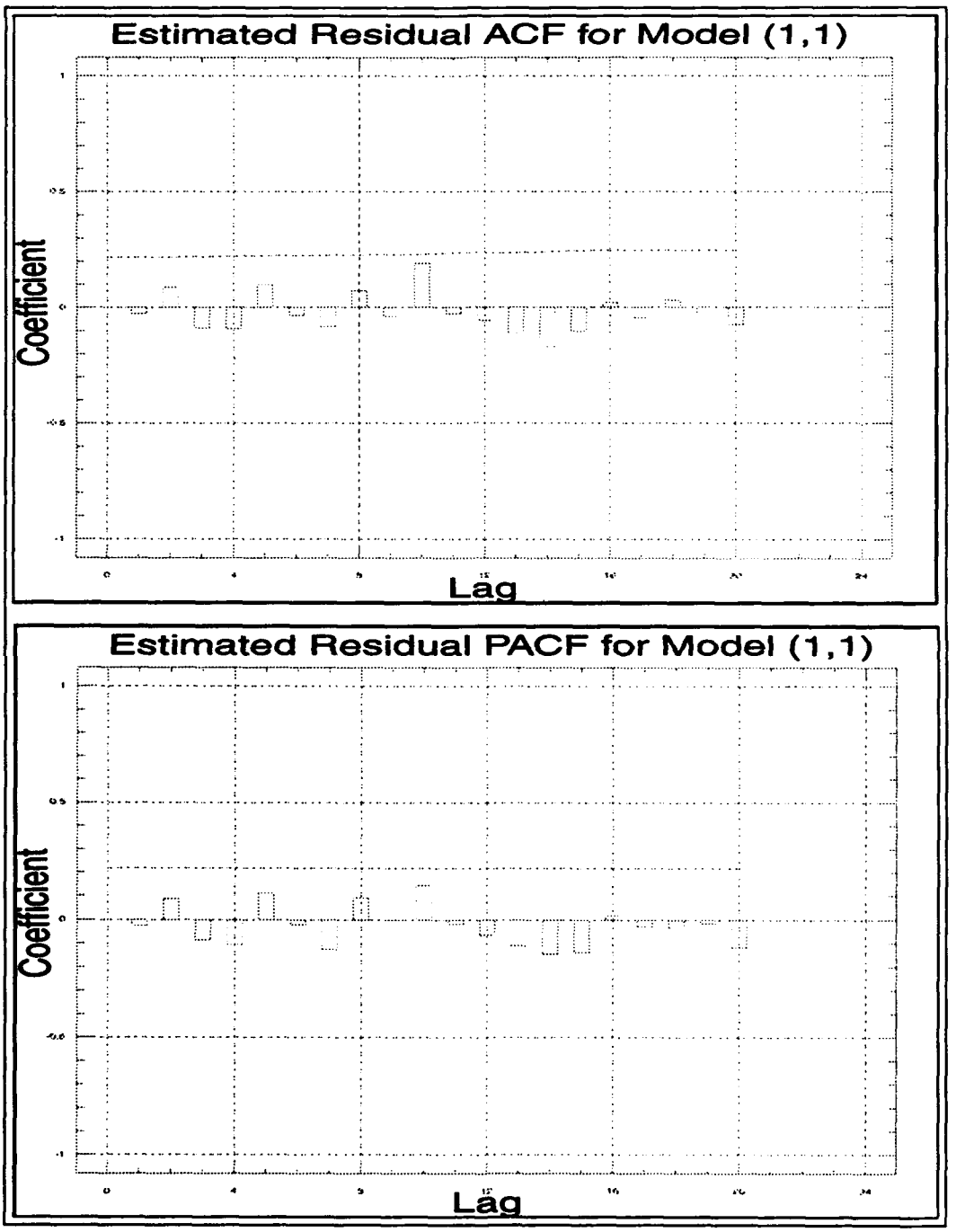

Figure F.36. Plots of the residual ACF and PACF for Model CS(1,0,1).

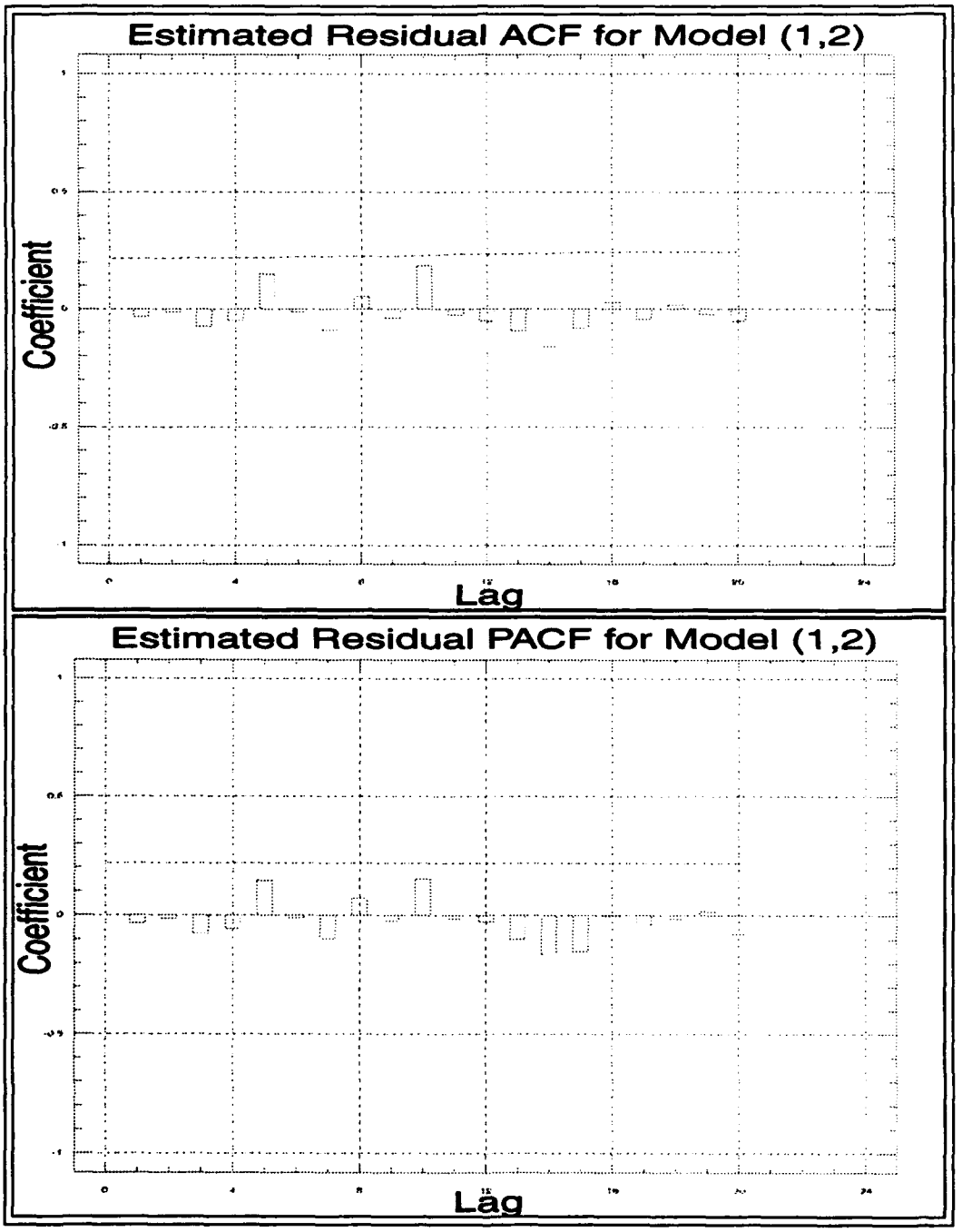

Figure F.37. Plots of the residual ACF and PACF for Model CS(1,0,2).

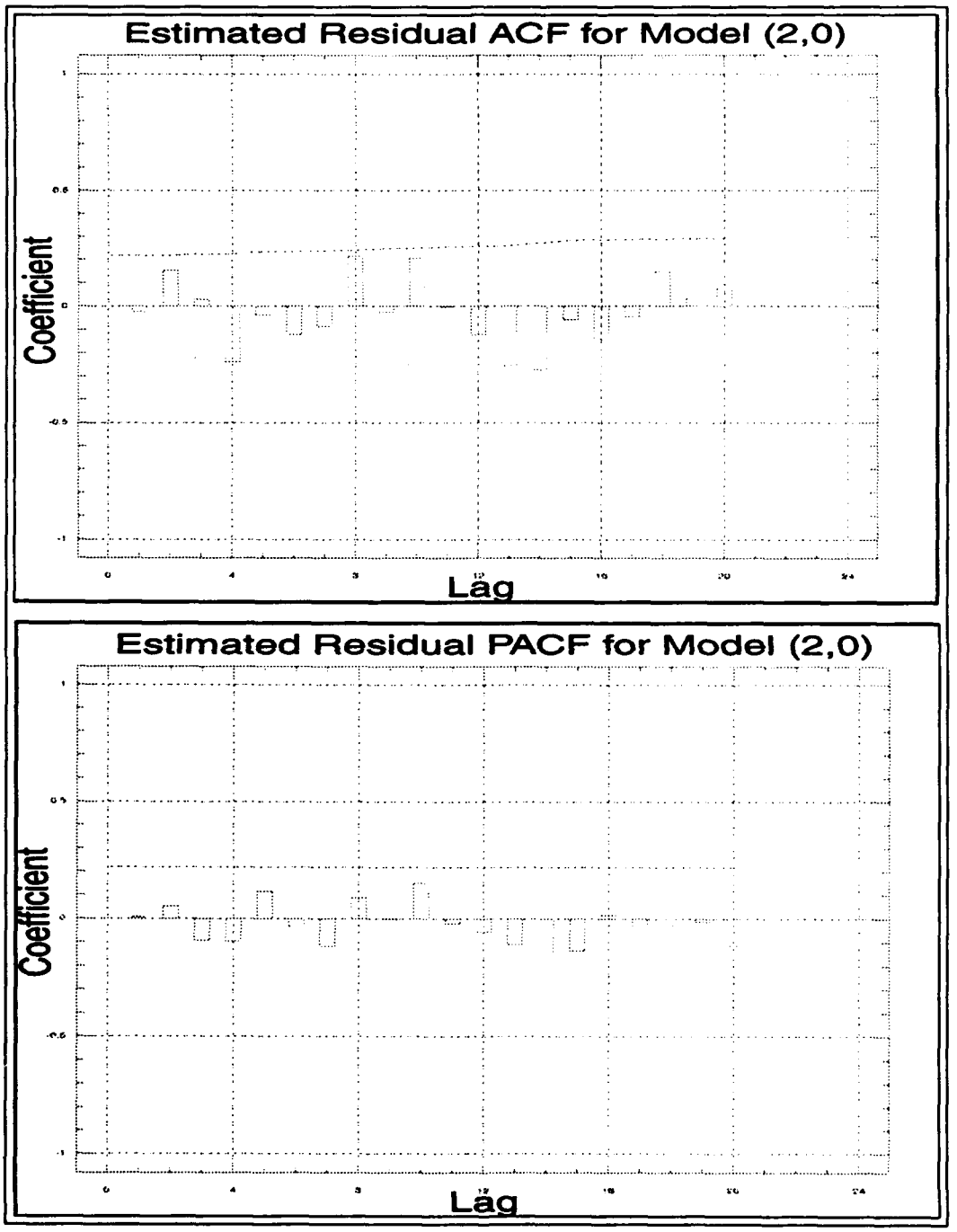

Figure F.38. Plots of the residual ACF and PACF for Model CS(2,0,0).

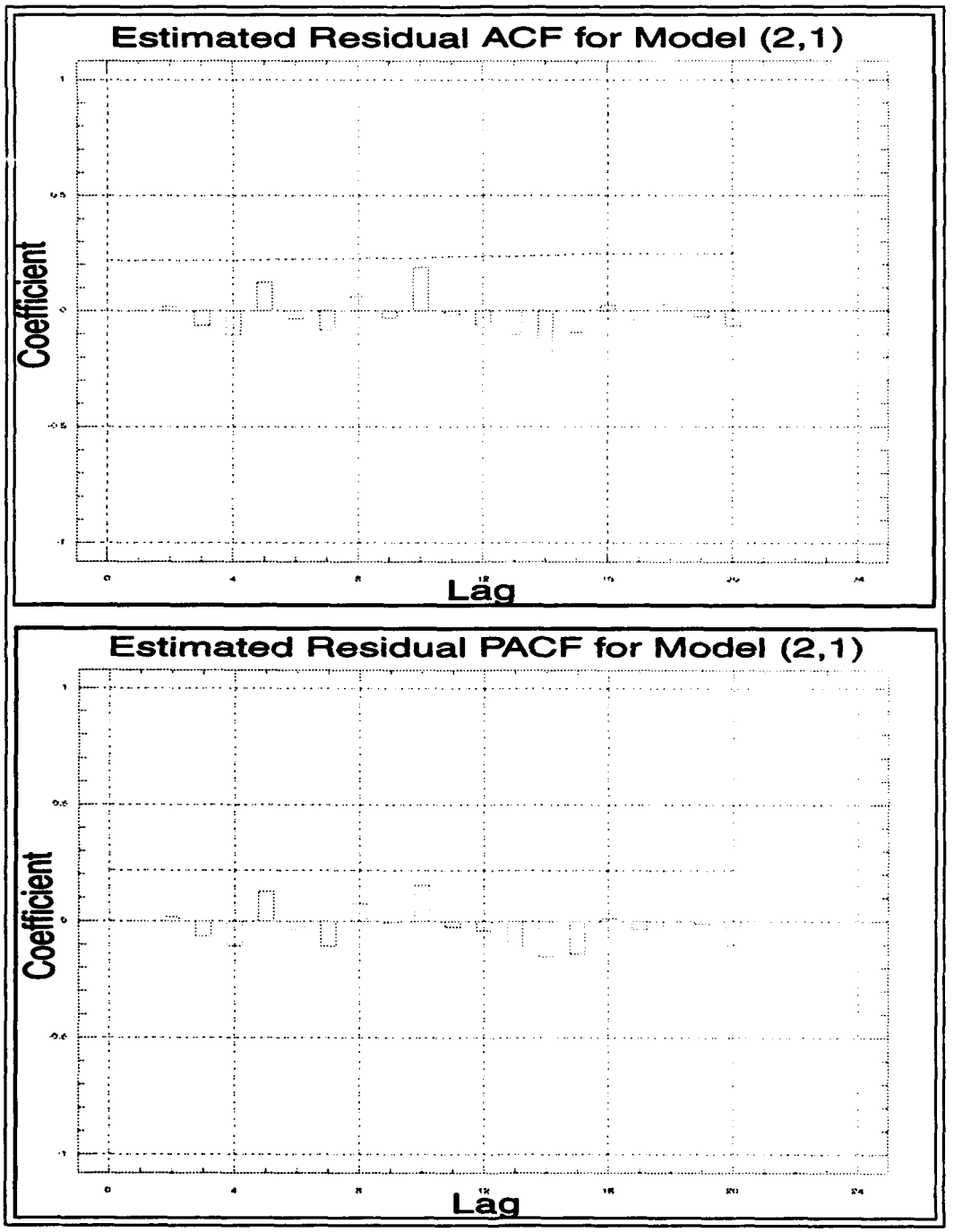

Figure F.39. Plots of the residual ACF and PACF for Model CS(2,0,1).

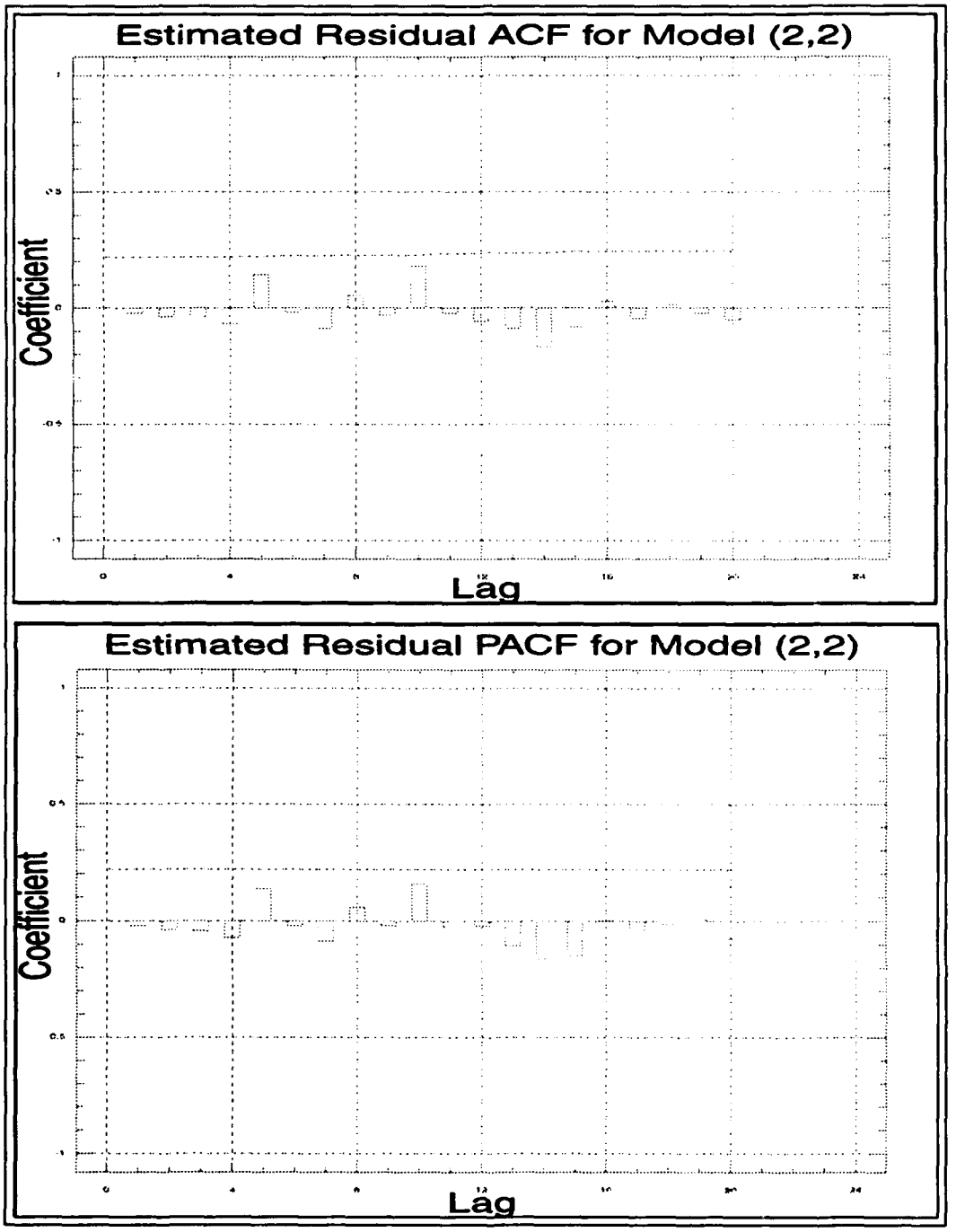

Figure F.40. Plots of the residual ACF and PACF for Model CS(2,0,2).

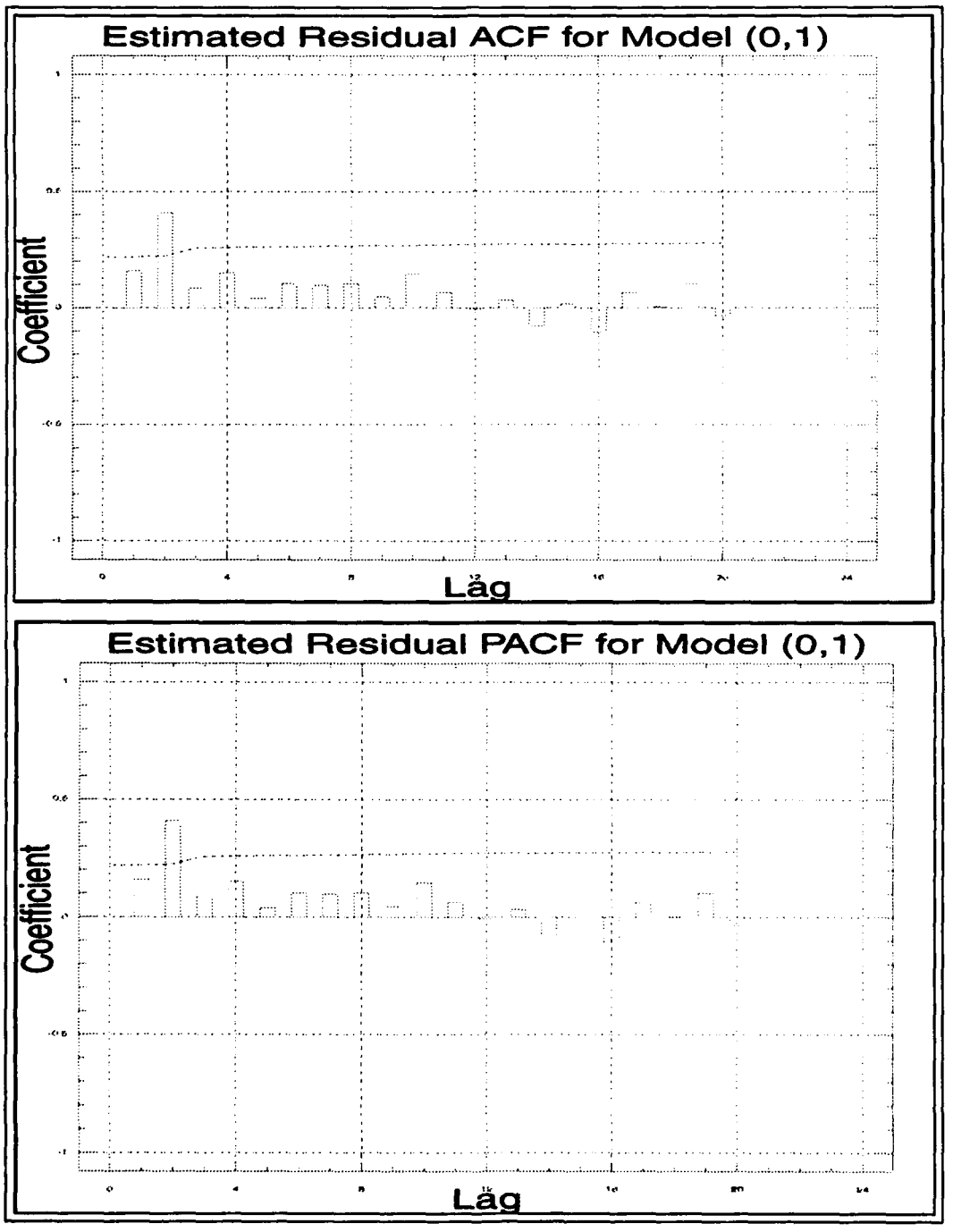

Figure F.41. Plots of the residual ACF and PACF for Model CW(0,0,1).

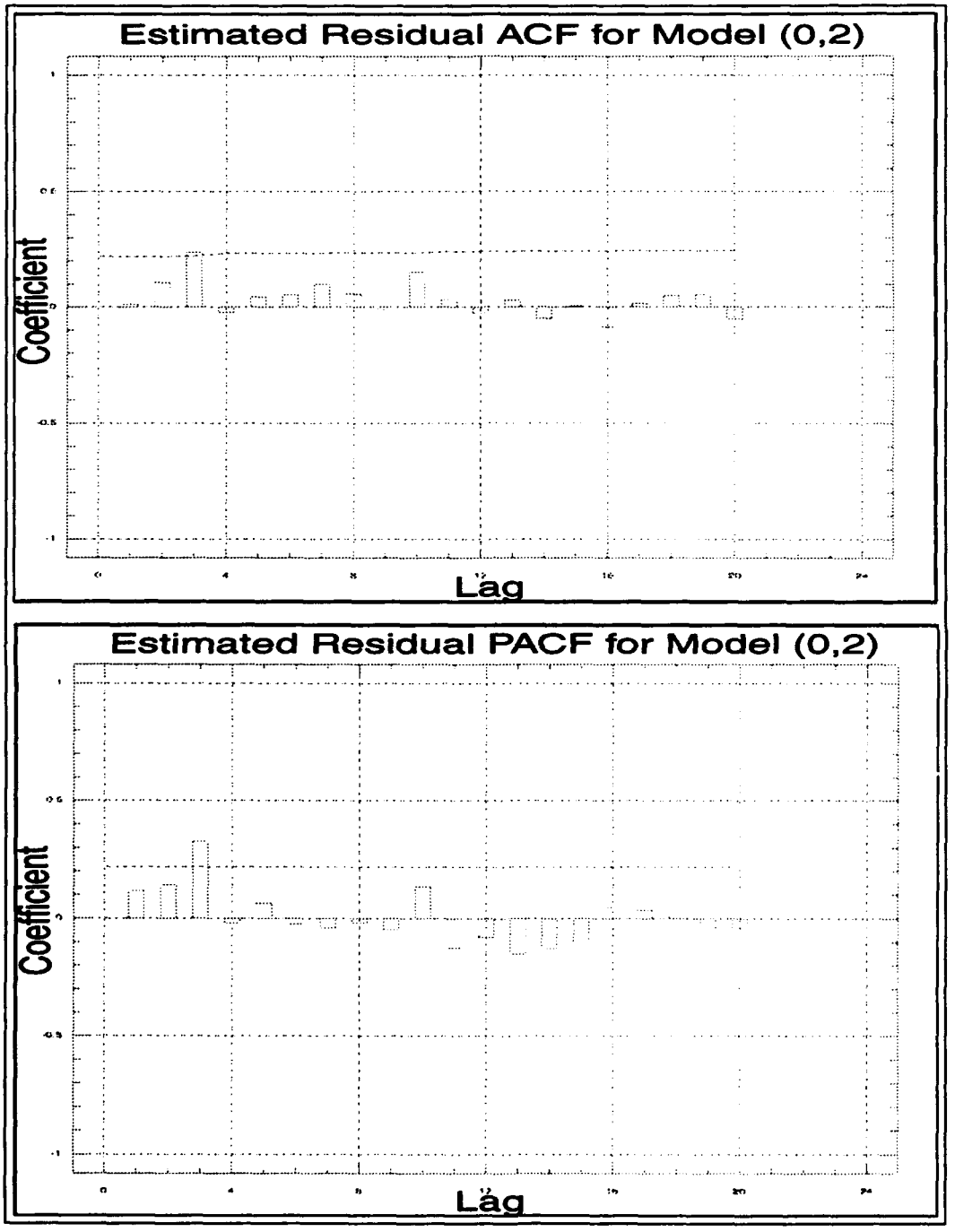

Figure F.42. Plots of the residual ACF and PACF for Model CW(0,0,2).

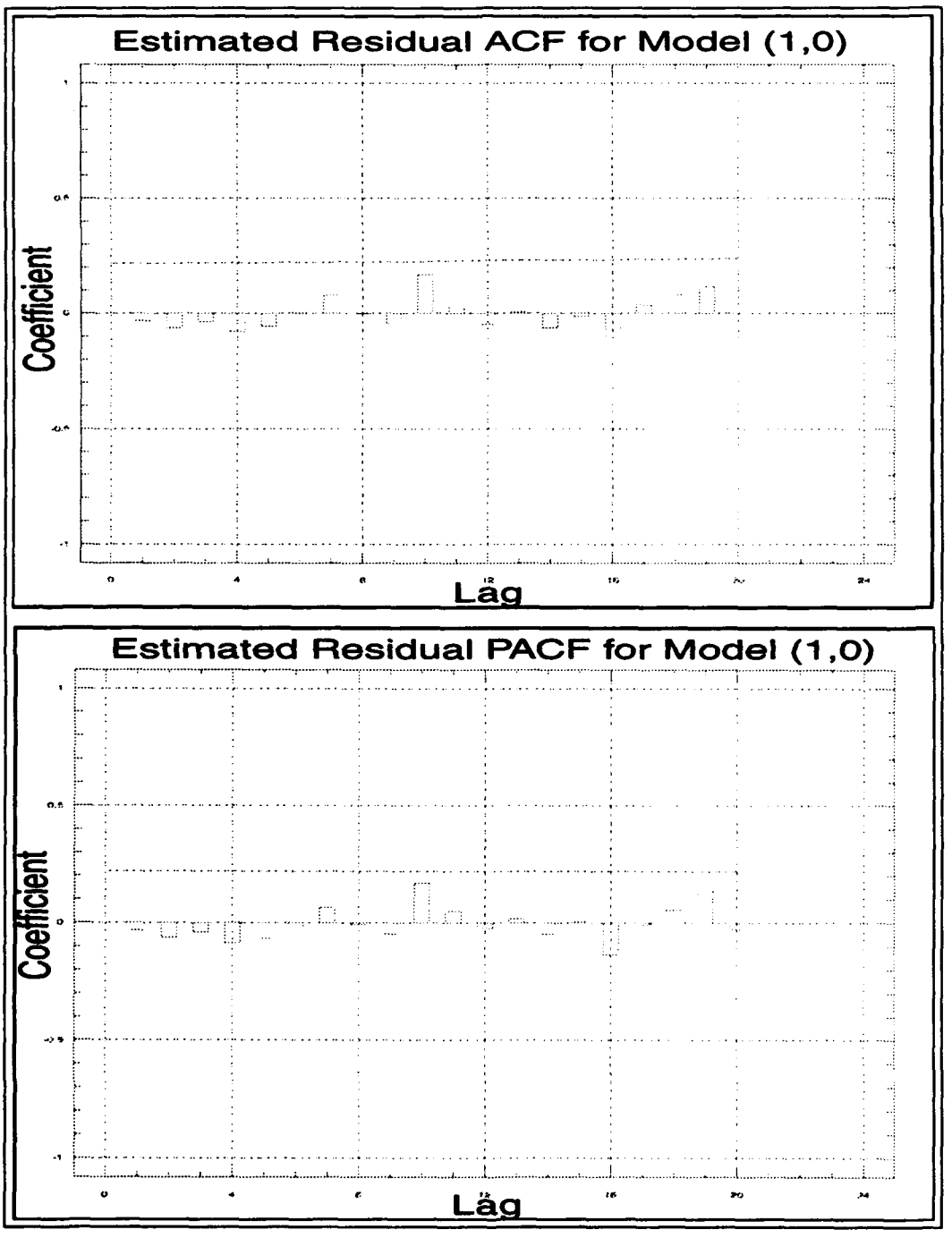

Figure F.43. Plots of the residual ACF and PACF for Model CW(1,0,0).

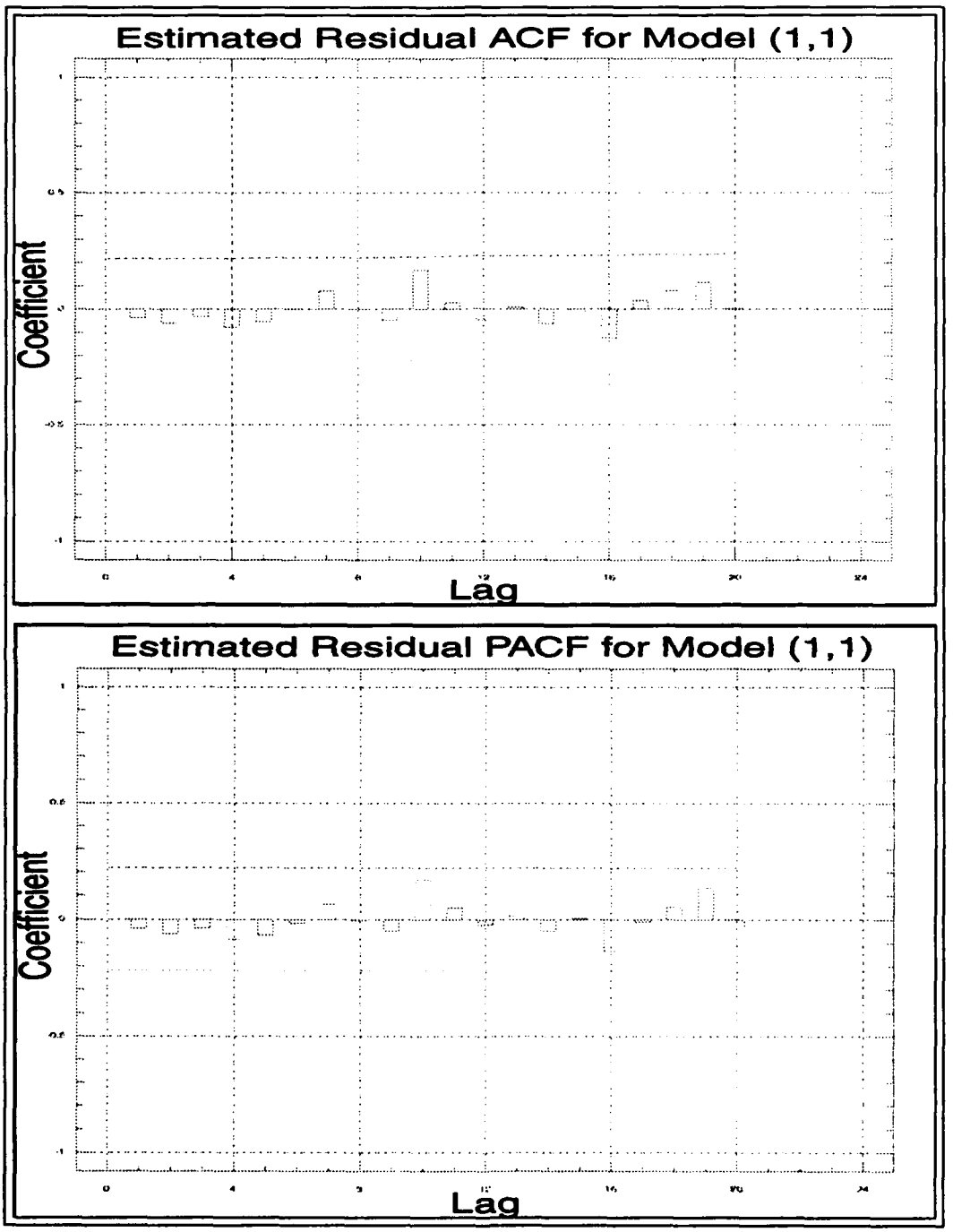

Figure F.44. Plots of the residual ACF and PACF for Model CW(1,0,1).

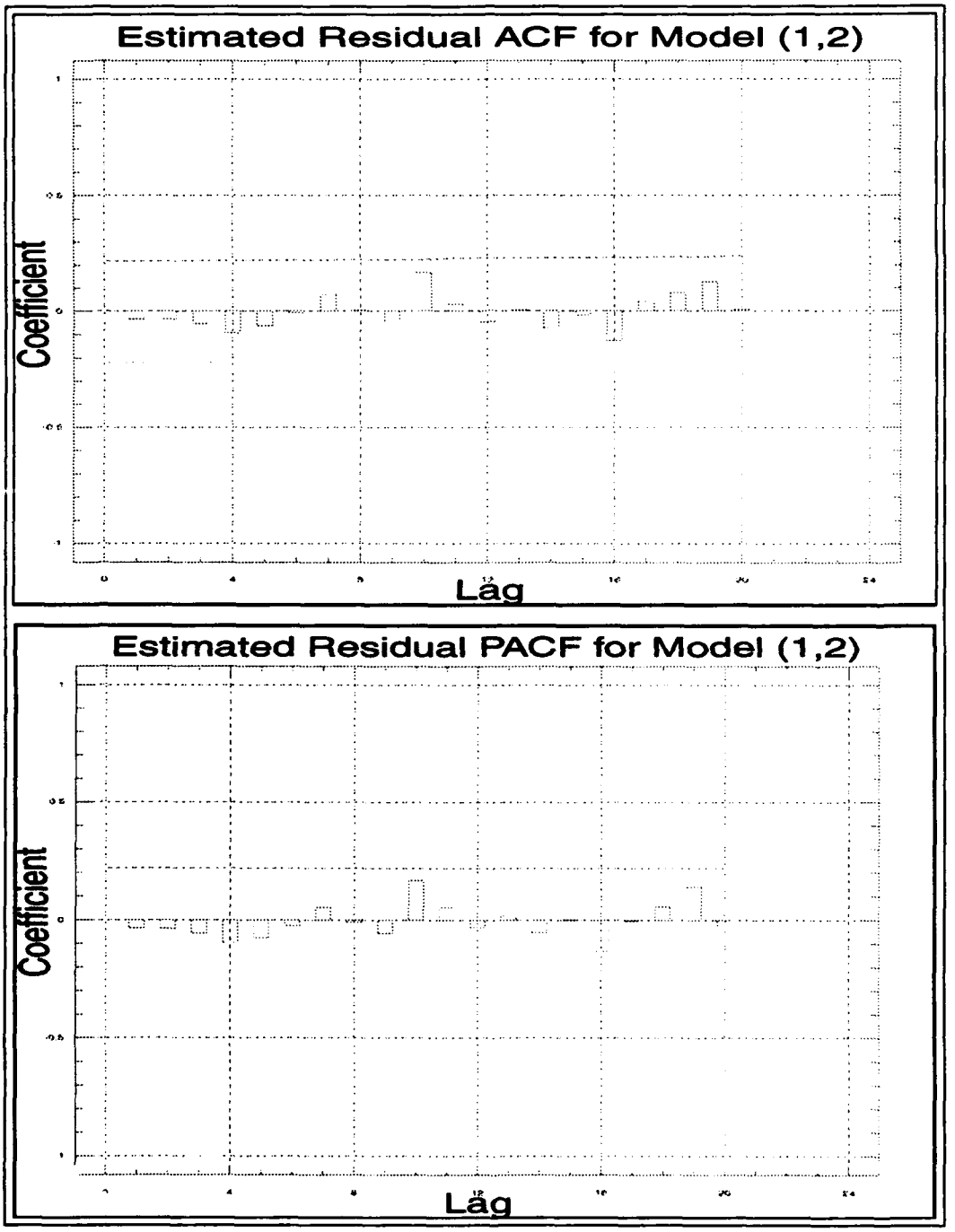

Figure F.45. Plots of the residual ACF and PACF for Model CW(1,0,2).

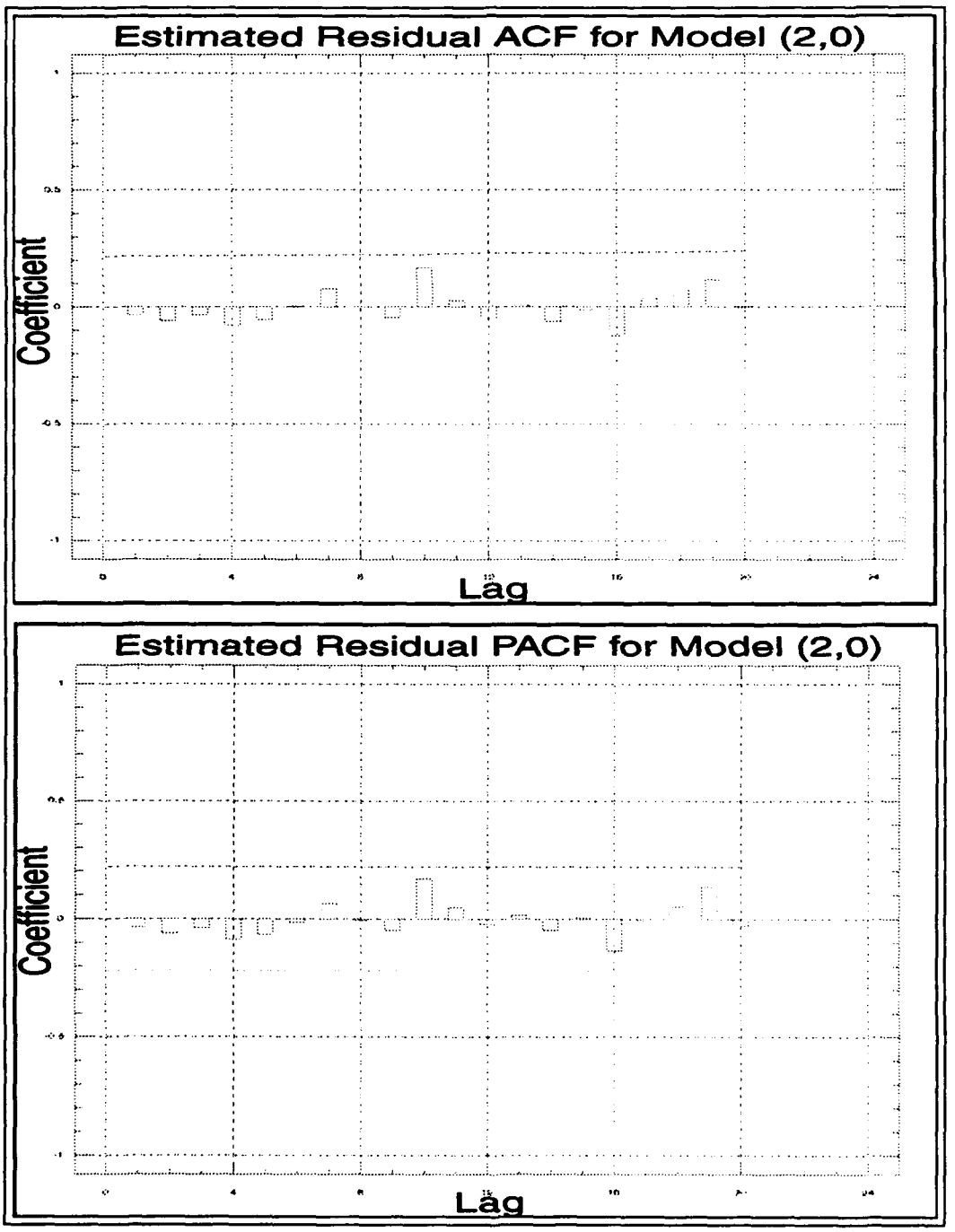

Figure F.46. Plots of the residual ACF and PACF for Model CW(2,0,0).

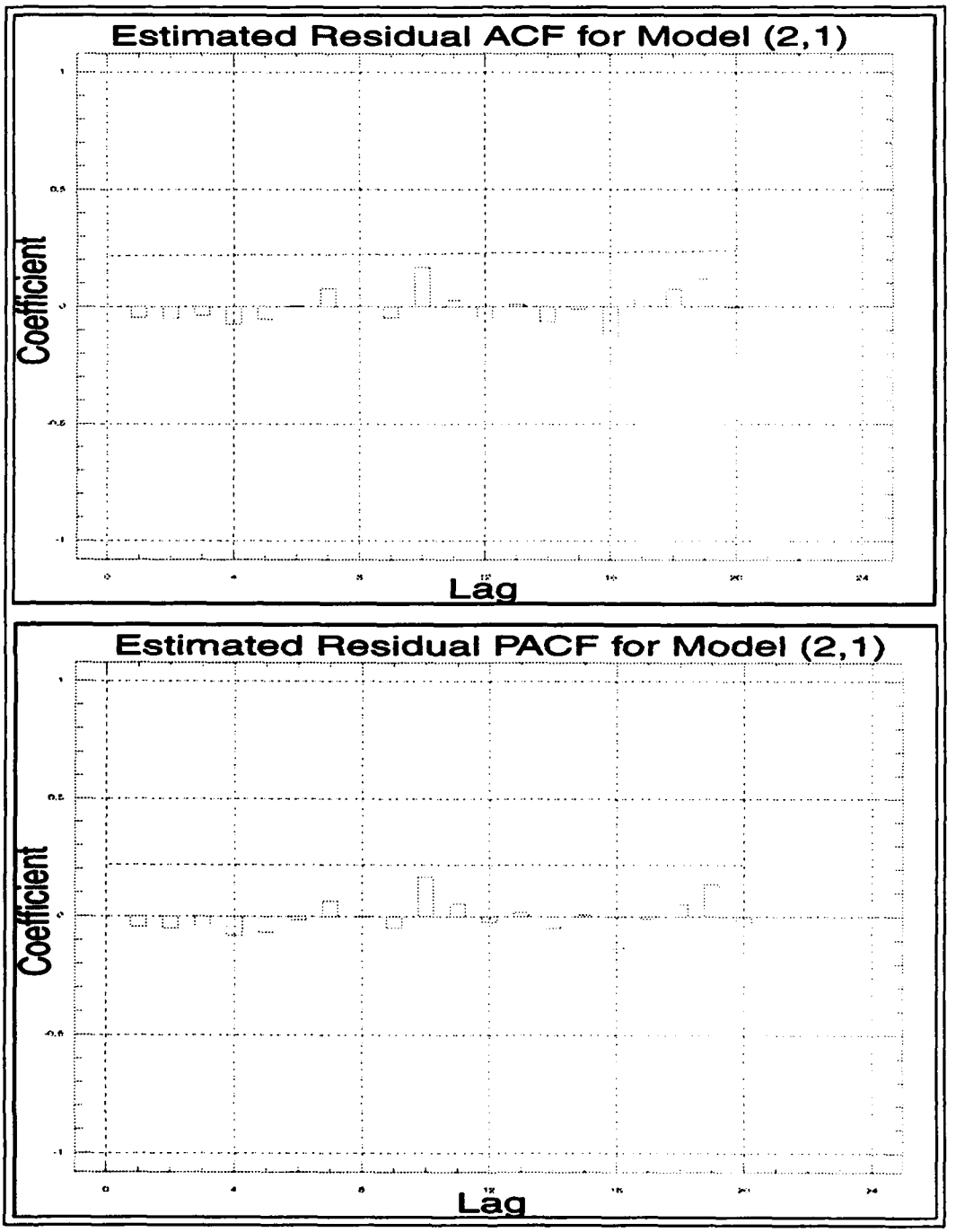

Figure F.47. Plots of the residual ACF and PACF for Model CW(2,0,1).

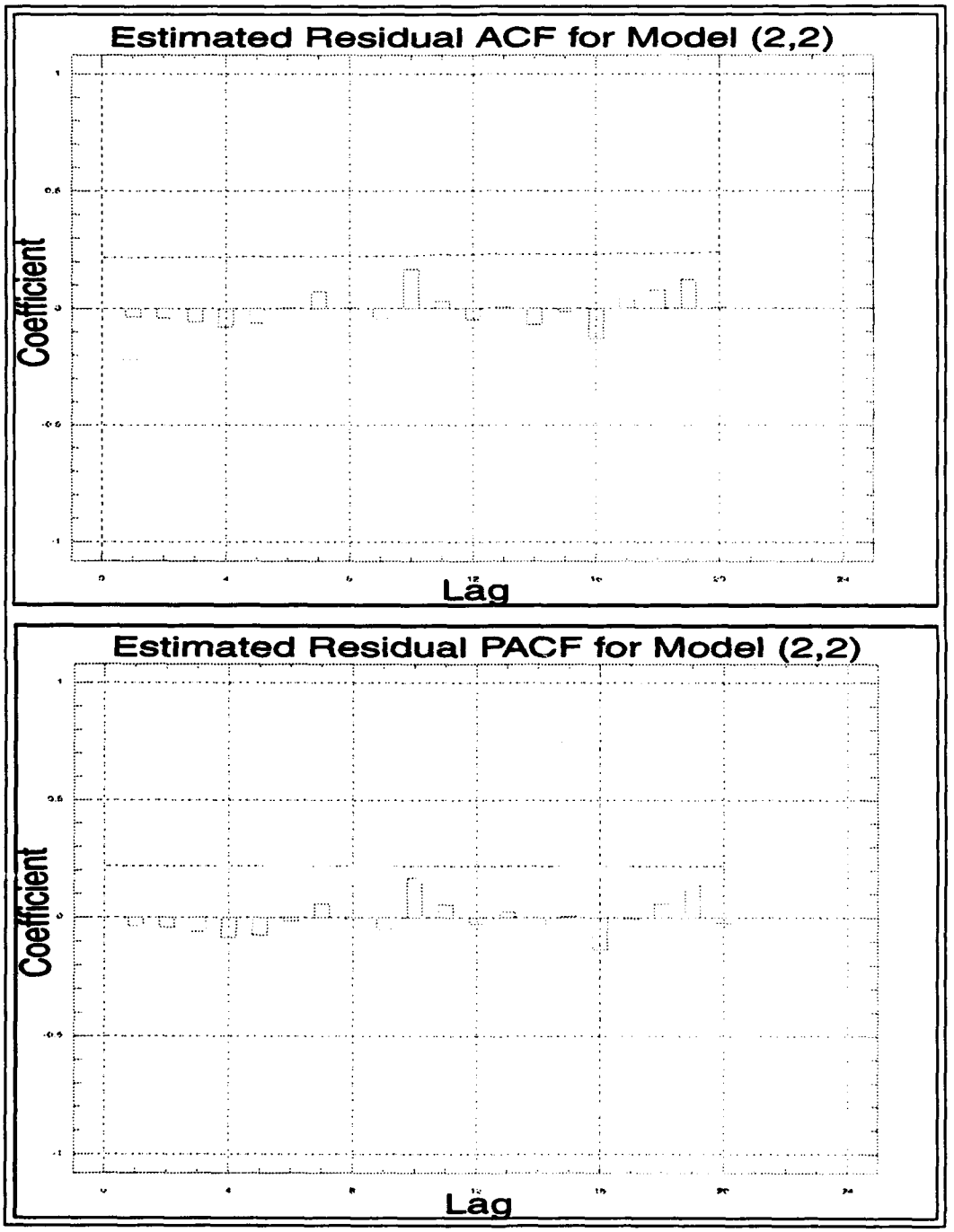

Figure F.48. Plots of the residual ACF and PACF for Model CW(2,0,2).

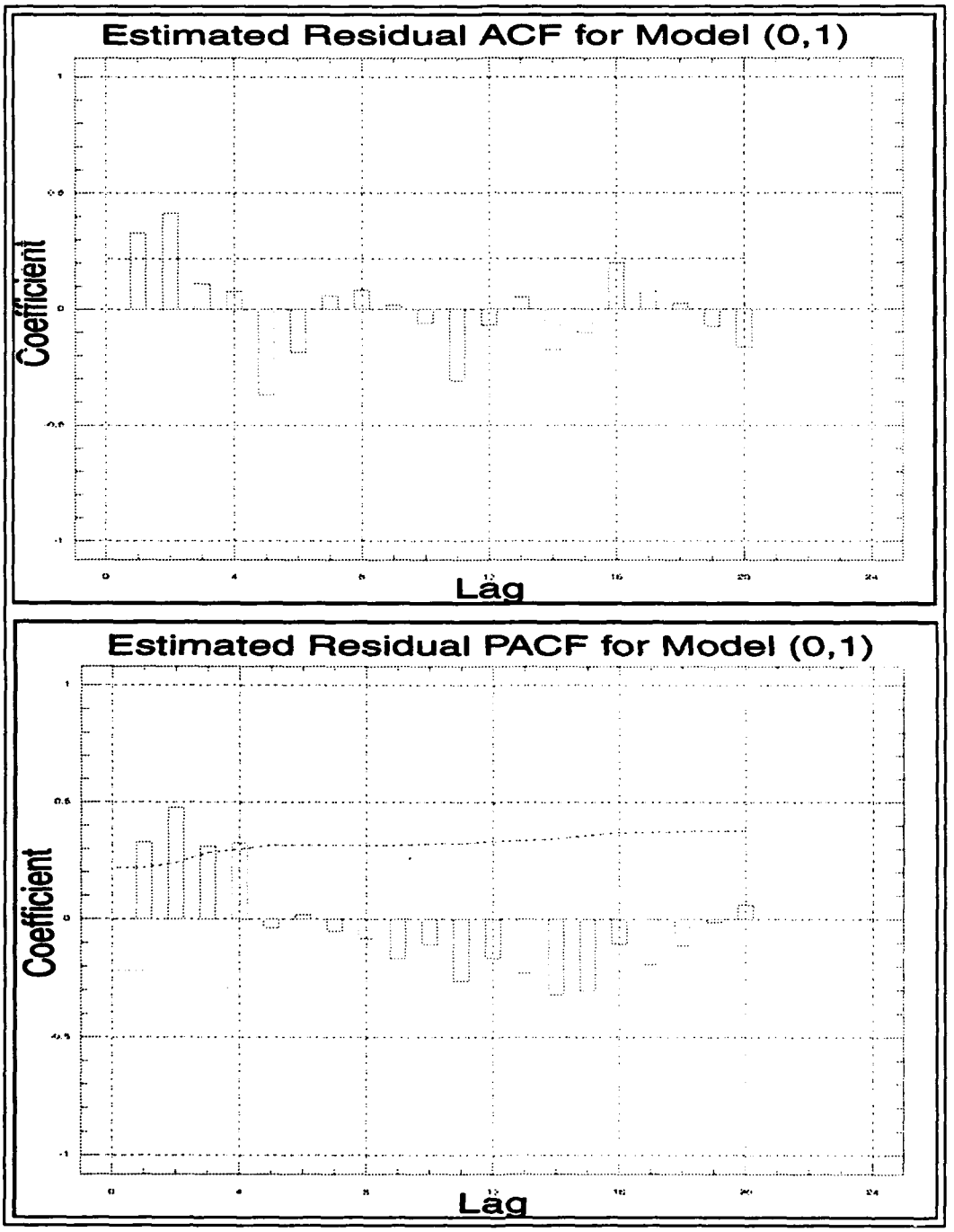

Figure F.49. Plots of the residual ACF and PACF for Model KN(0,0,1).

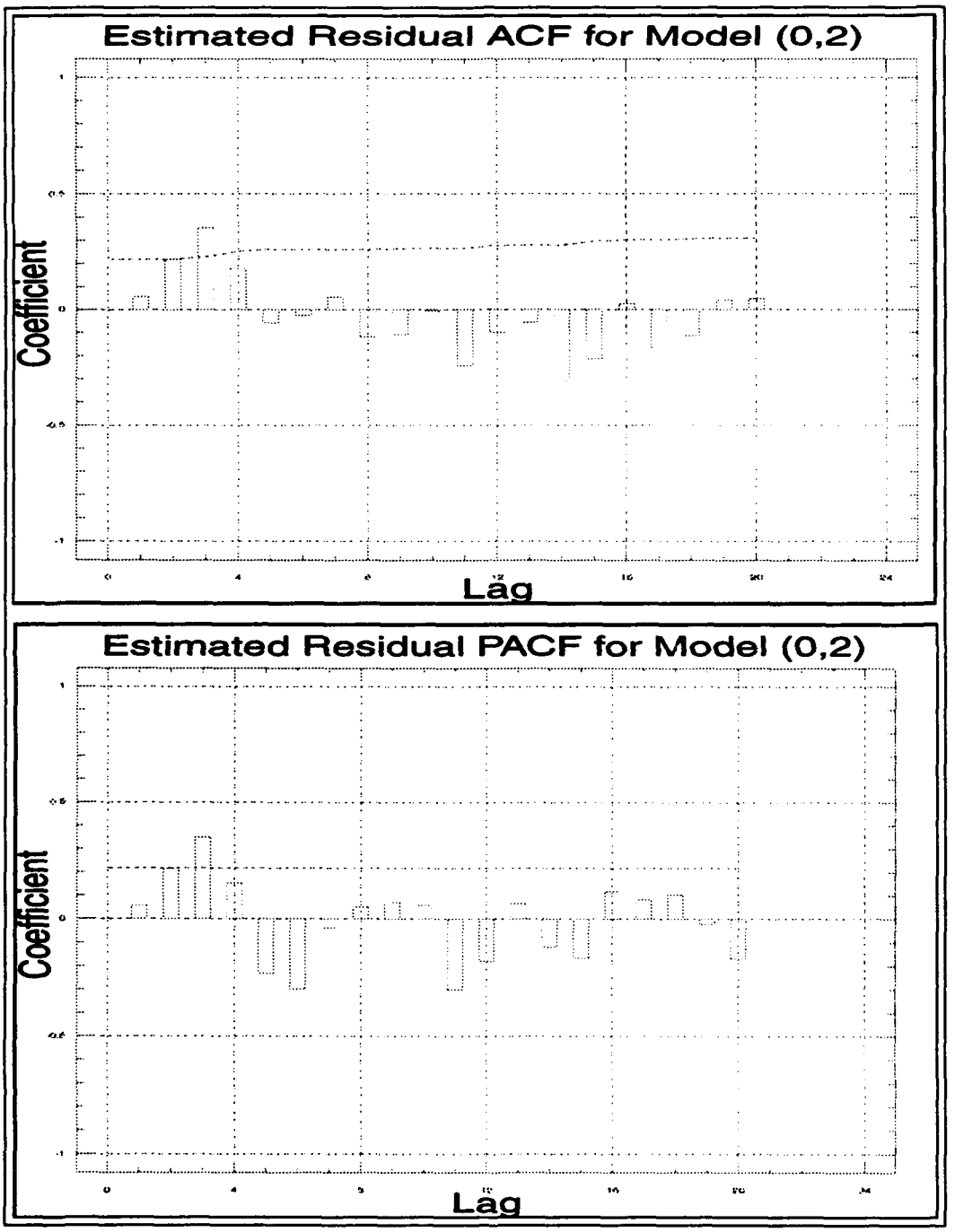

Figure F.50. Plots of the residual ACF and PACF for Model KN(0,0,2).

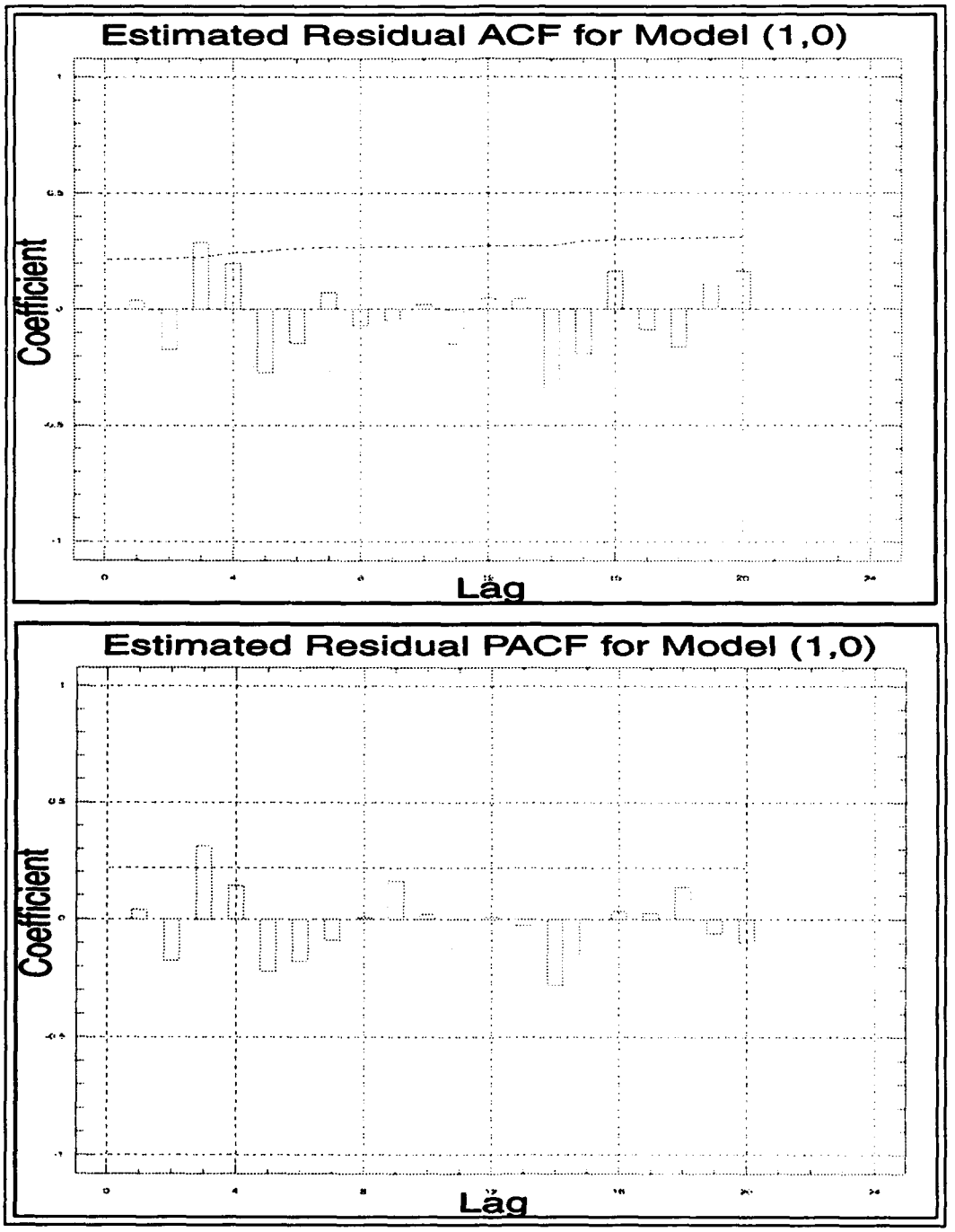

Figure F.51. Plots of the residual ACF and PACF for Model KN(1,0,0).

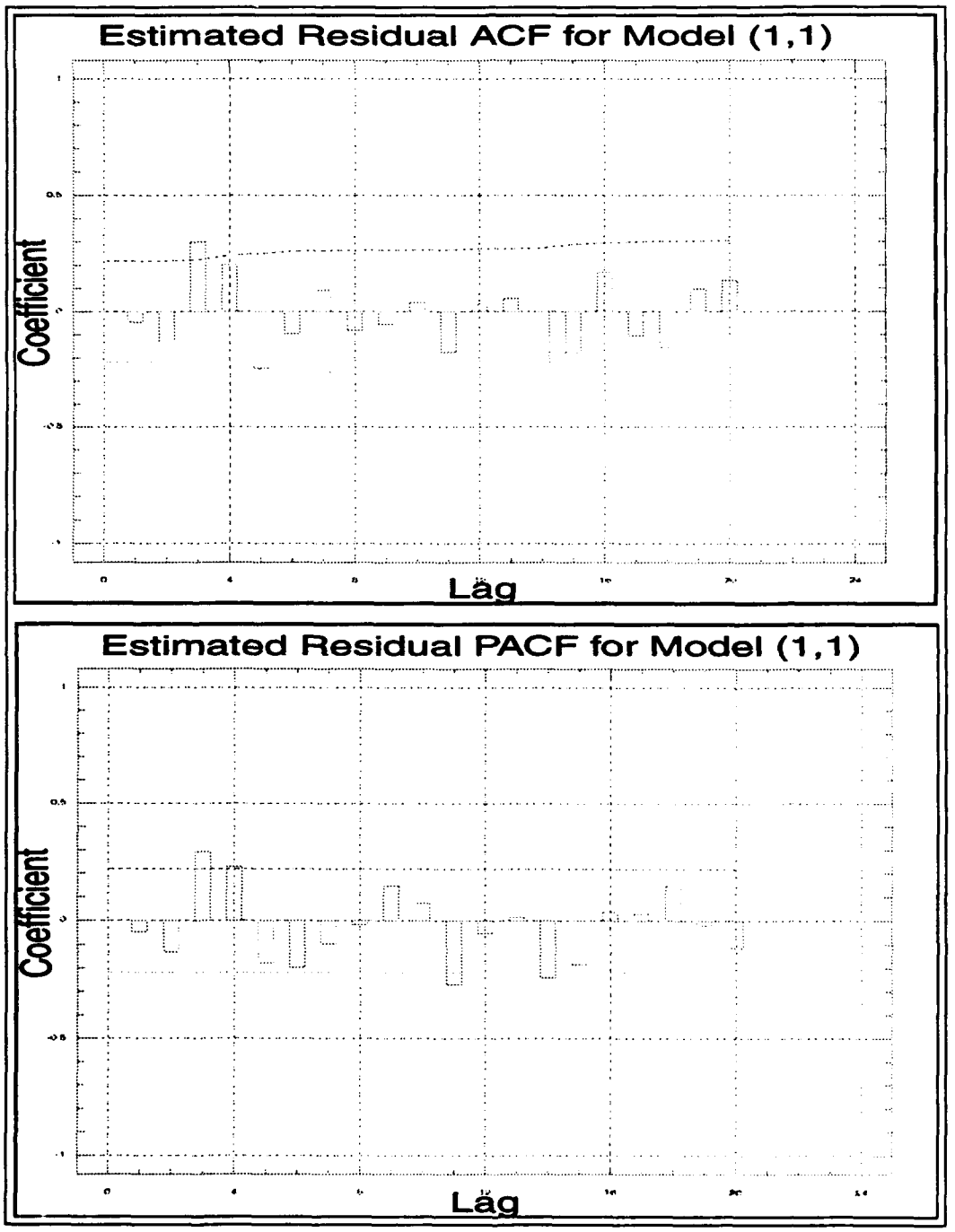

Figure F.52. Plots of the residual ACF and PACF for Model KN(1,0,1).

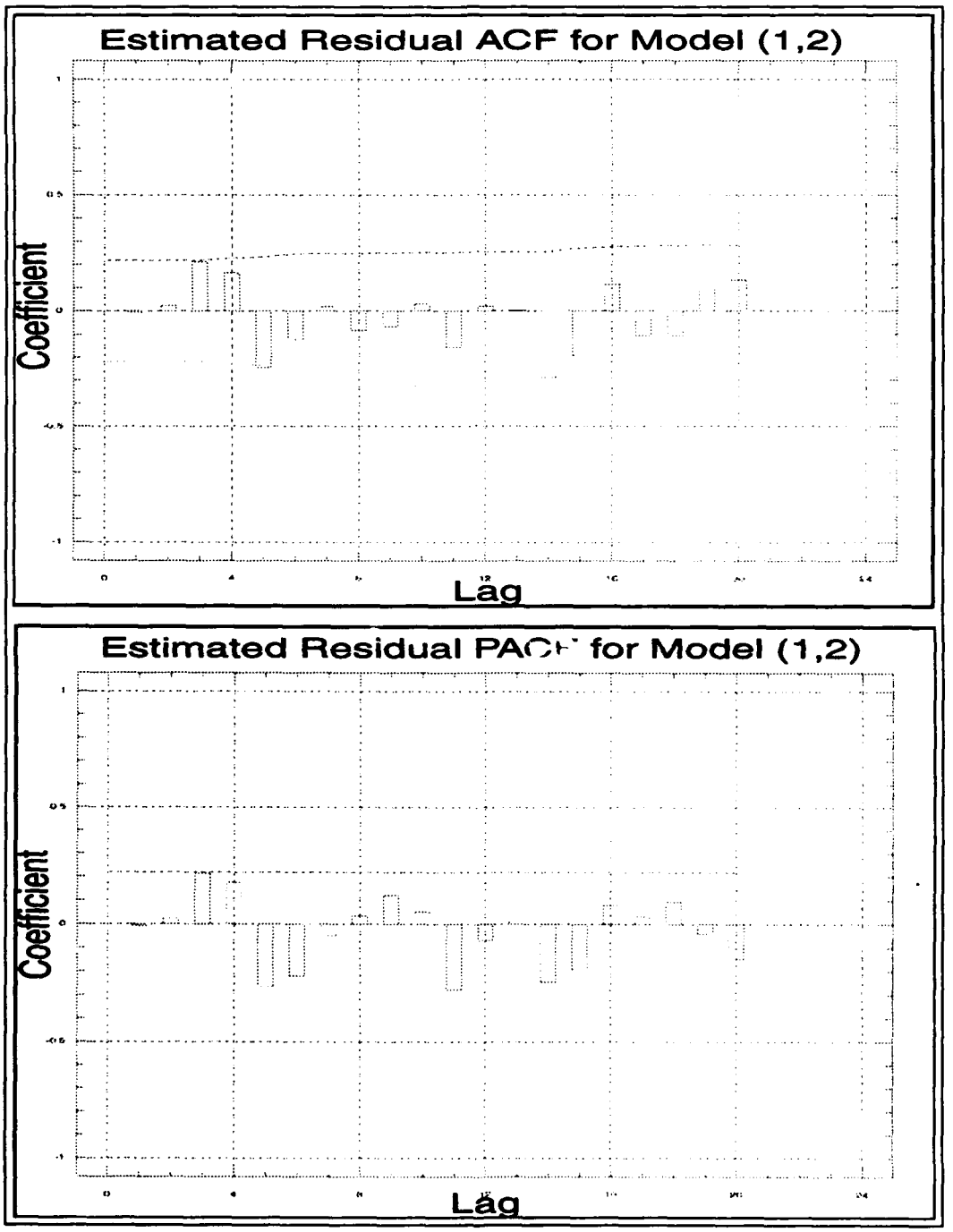

Figure F.53. Plots of the residual ACF and PACF for Model KN(1,0,2).

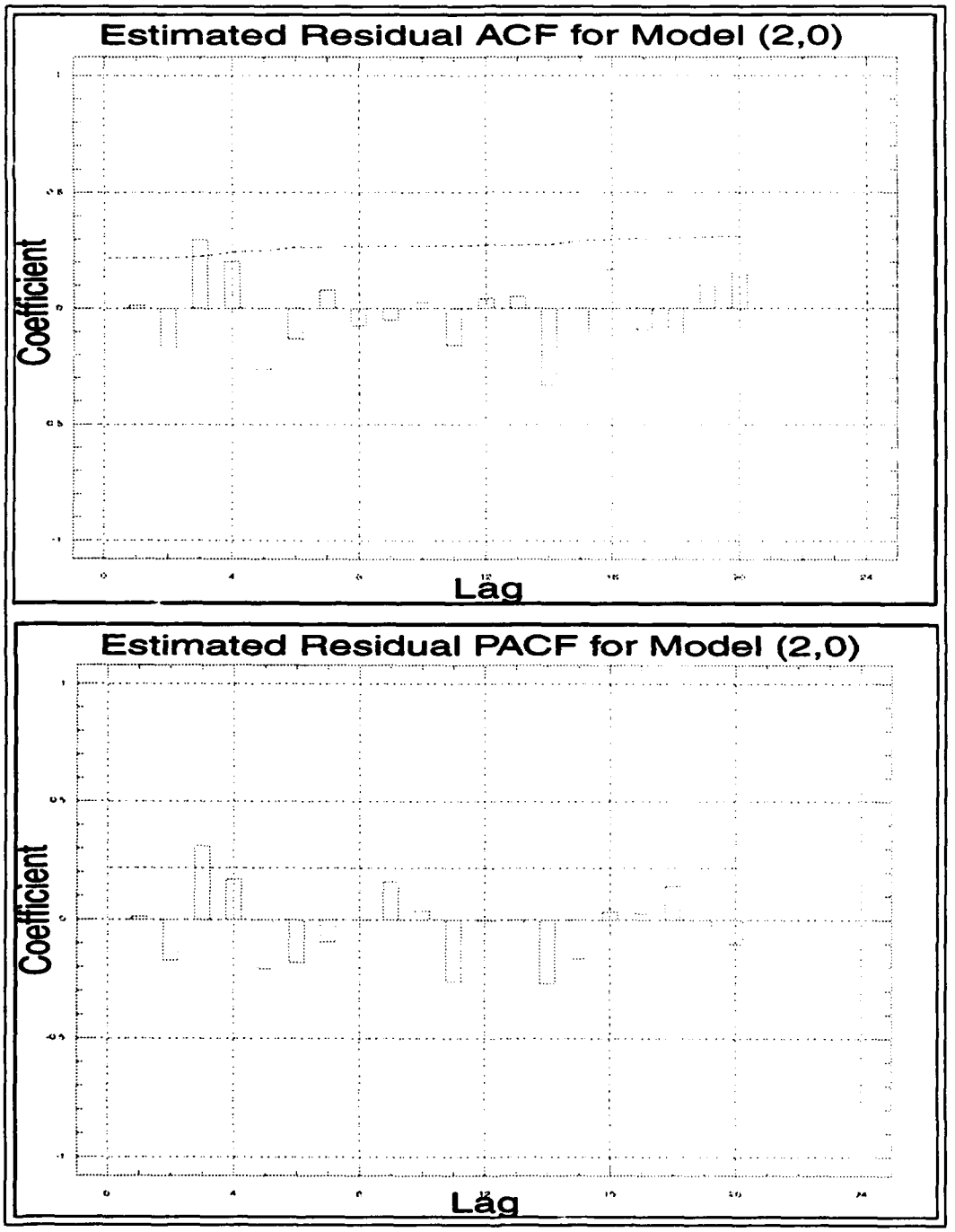

Figure F.54. Plots of the residual ACF and PACF for Model KN(2,0,0).

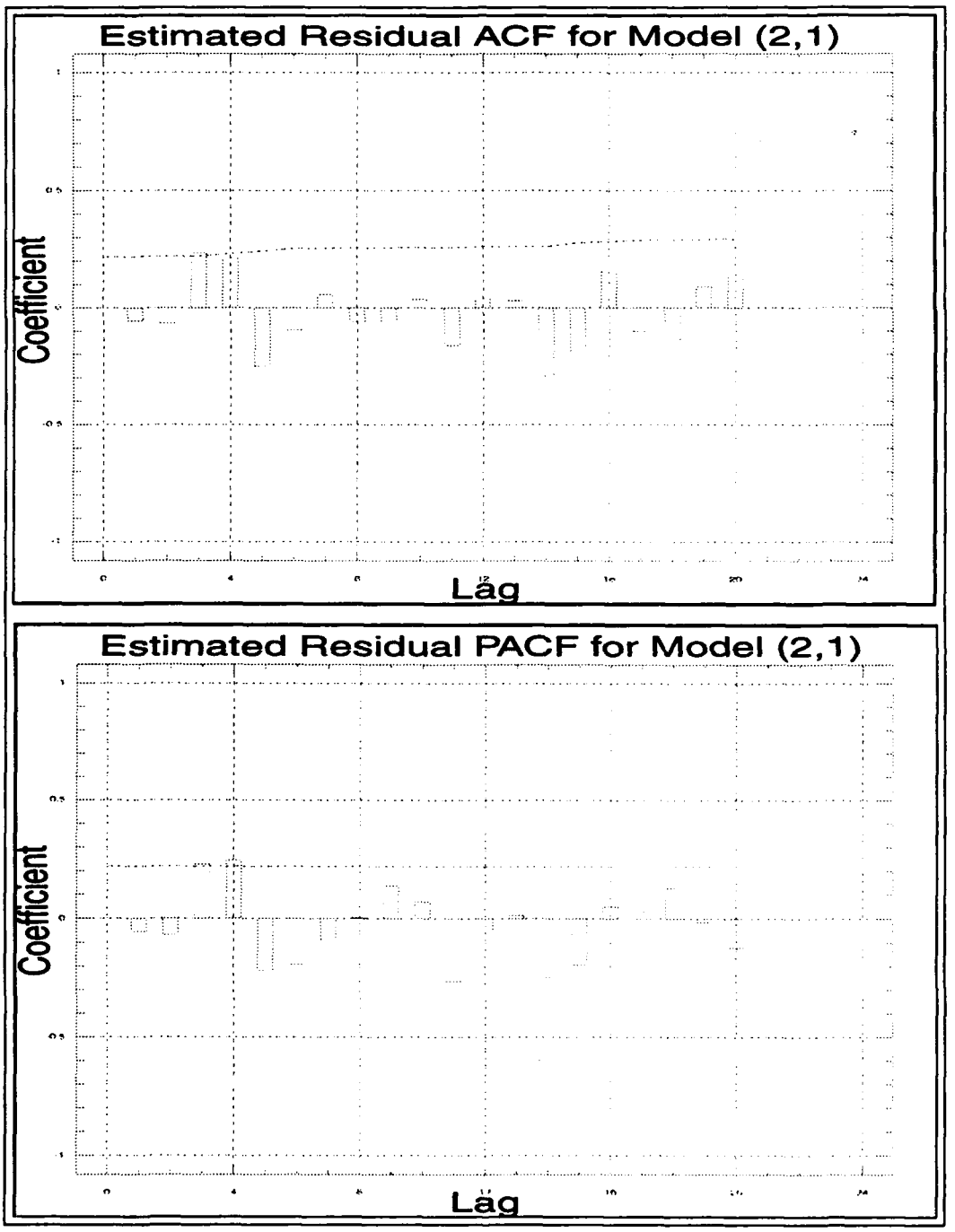

Figure F.55. Plots of the residual ACF and PACF for Model KN(2,0,1).

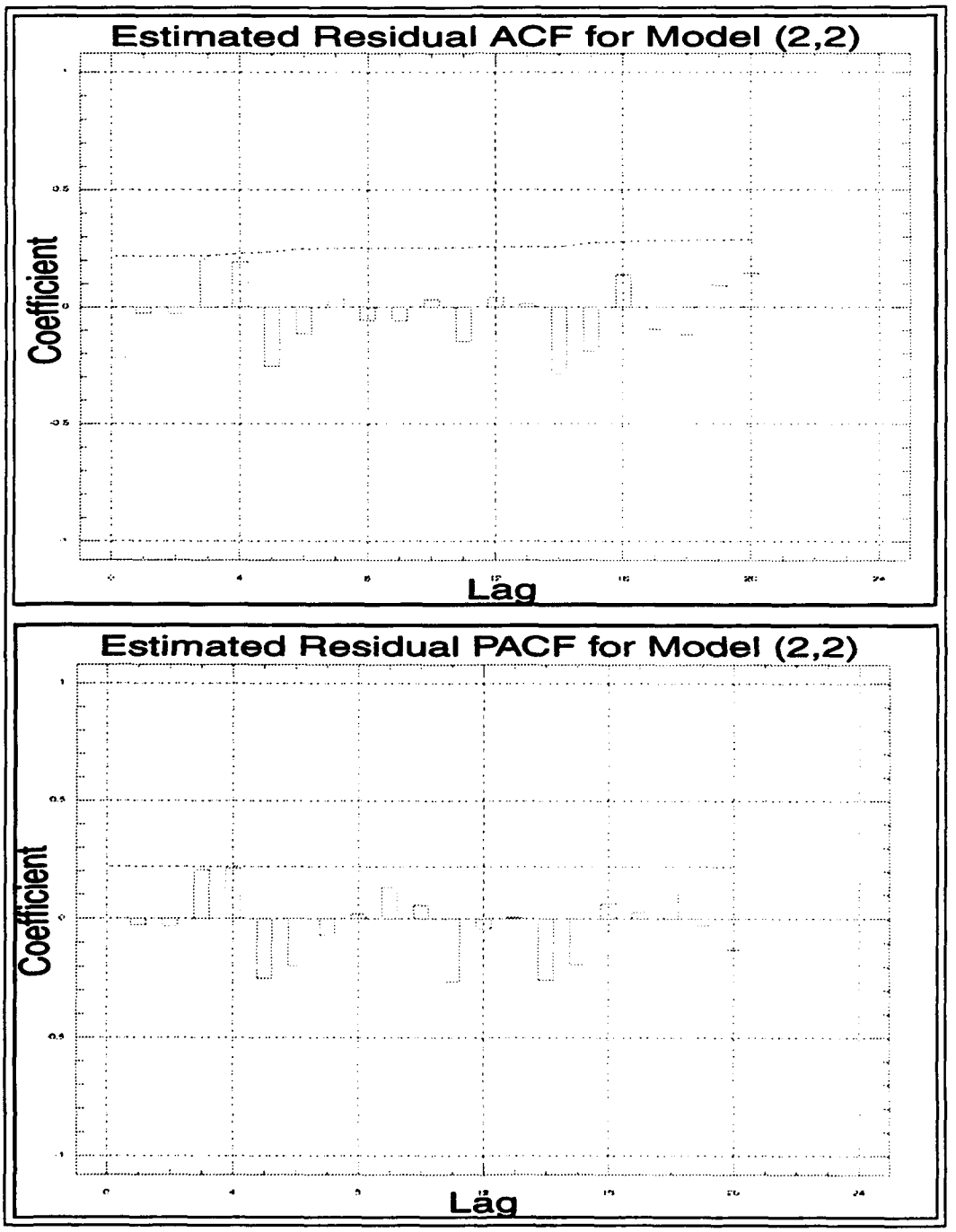

Figure F.56. Plots of the residual ACF and PACF for Model KN(2,0,2).

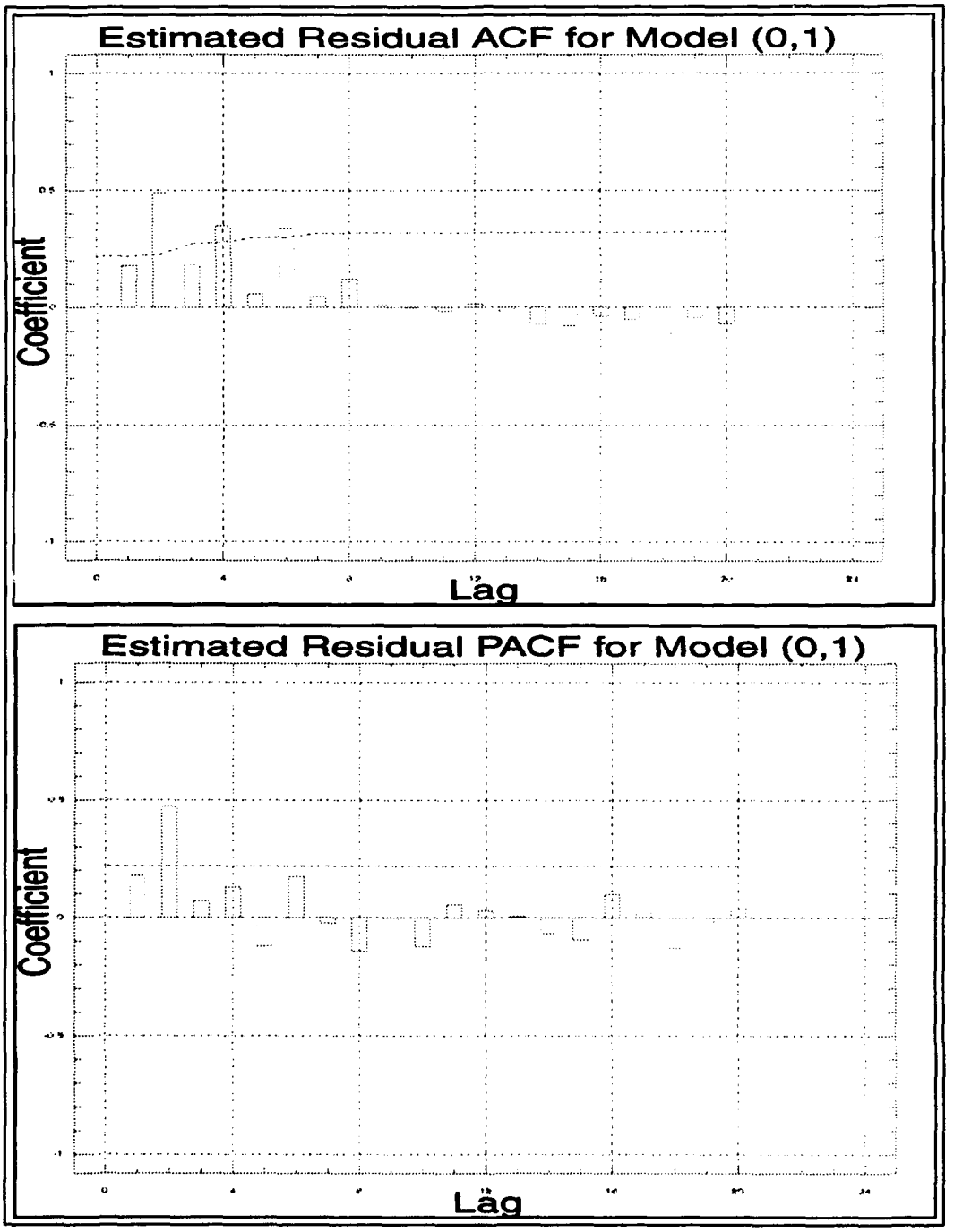

Figure F.57. Plots of the residual ACF and PACF for Model KE(0,0,1).

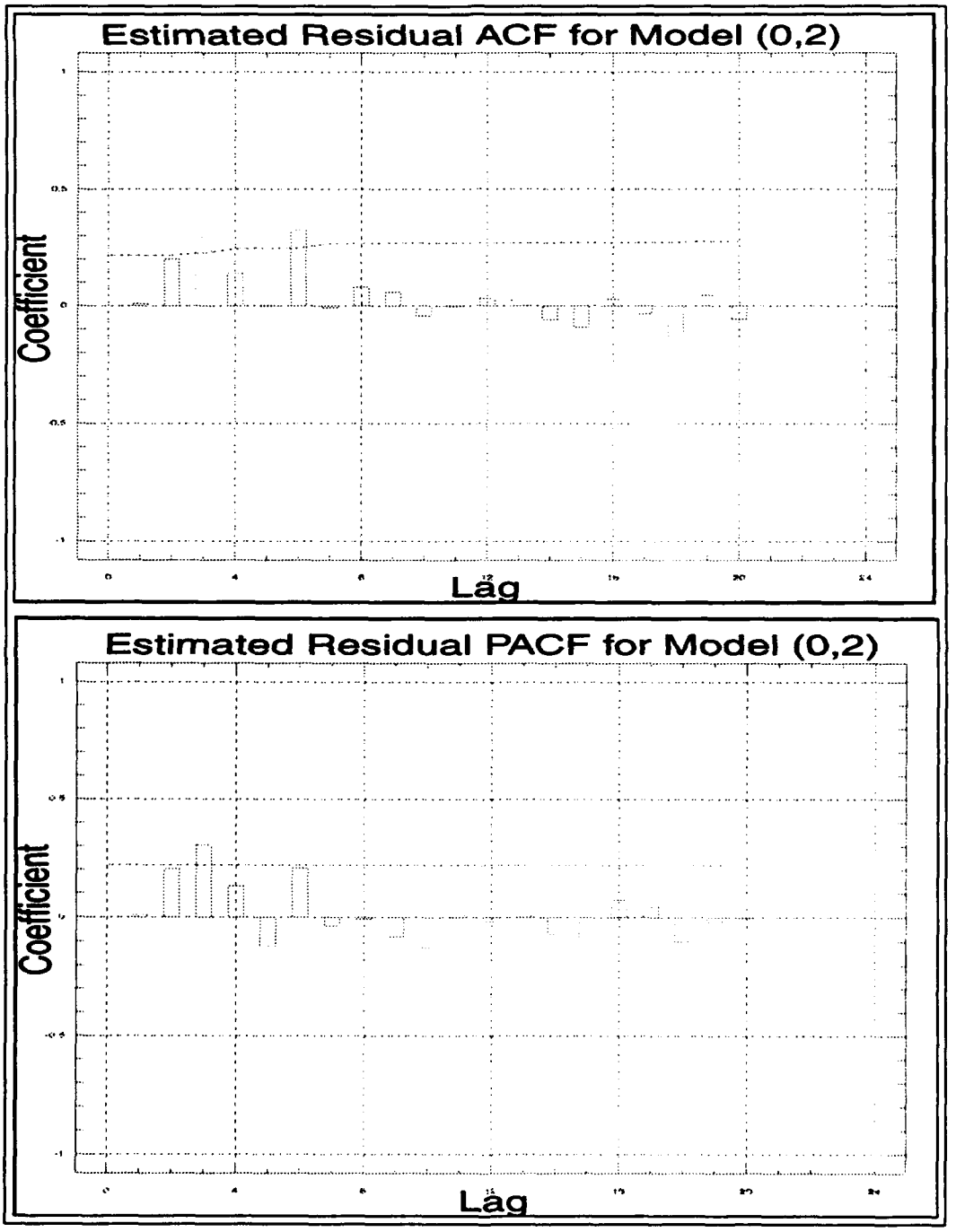

Figure F.58. Plots of the residual ACF and PACF for Model KE(0,0,2).

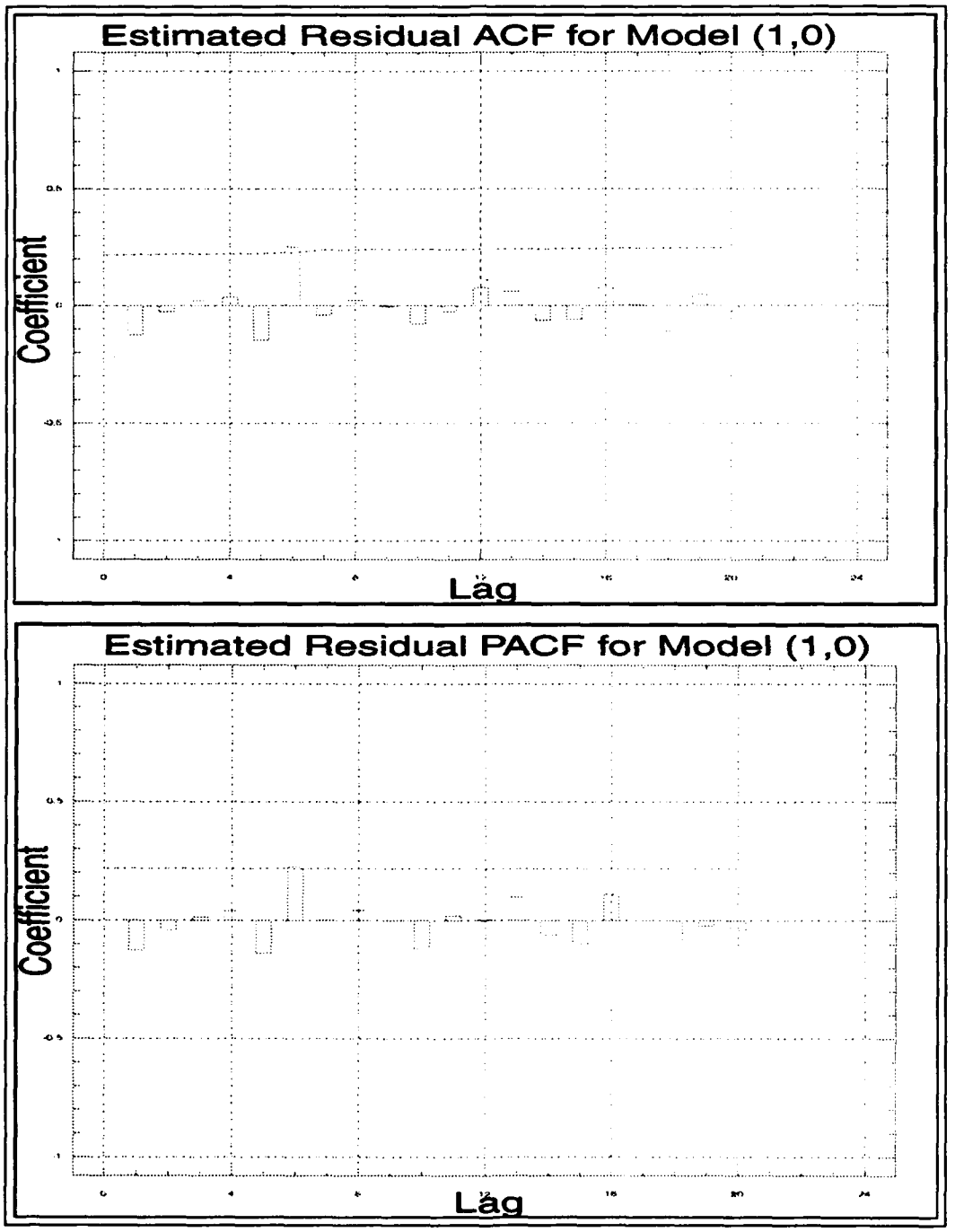

Figure F.59. Plots of the residual ACF and PACF for Model KE(1,0,0).

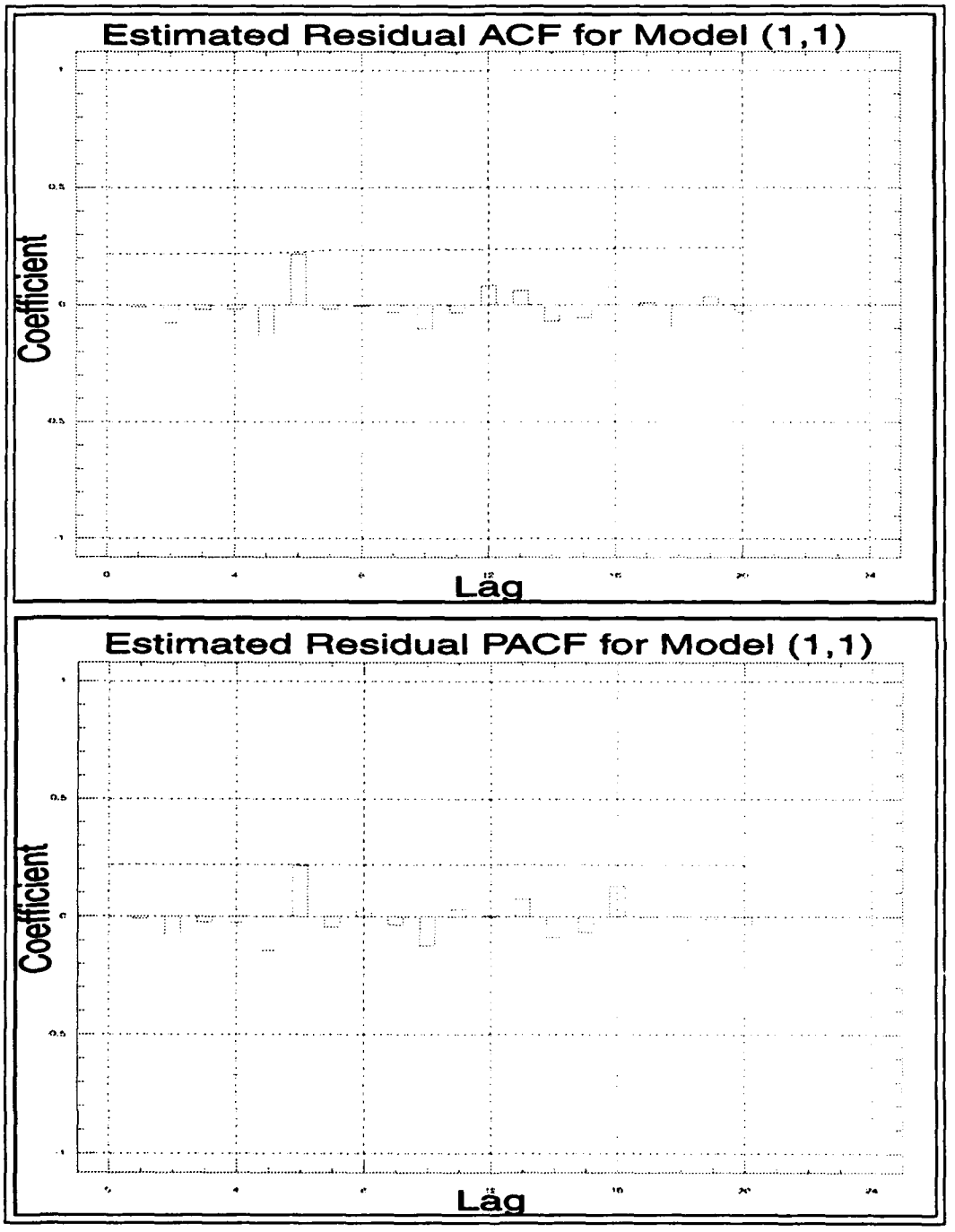

Figure F.60. Plots of the residual ACF and PACF for Model KE(1,0,1).

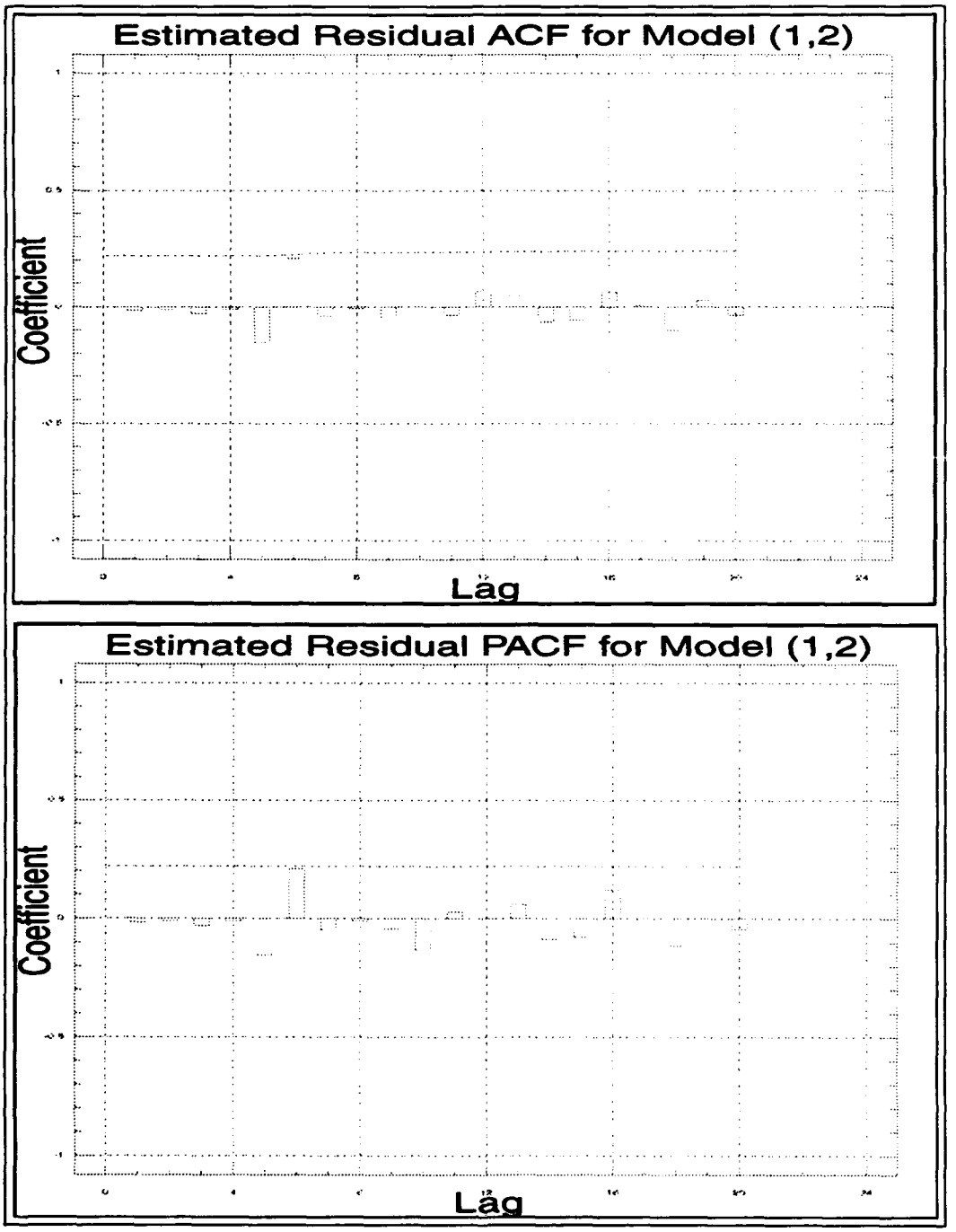

Figure F.61. Plots of the residual ACF and PACF for Model KE(1,0,2).

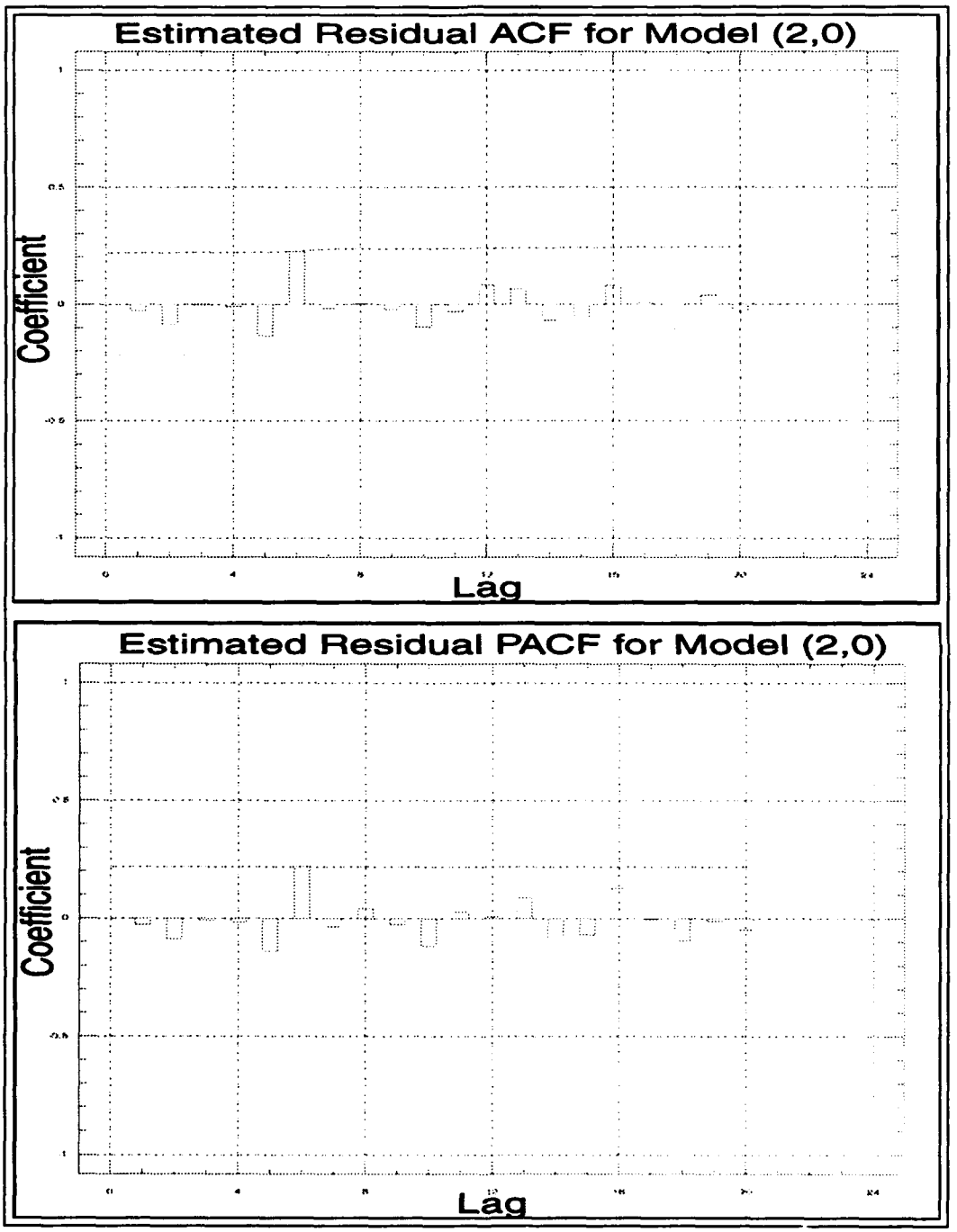

Figure F.62. Plots of the residual ACF and PACF for Model KE(2,0,0).

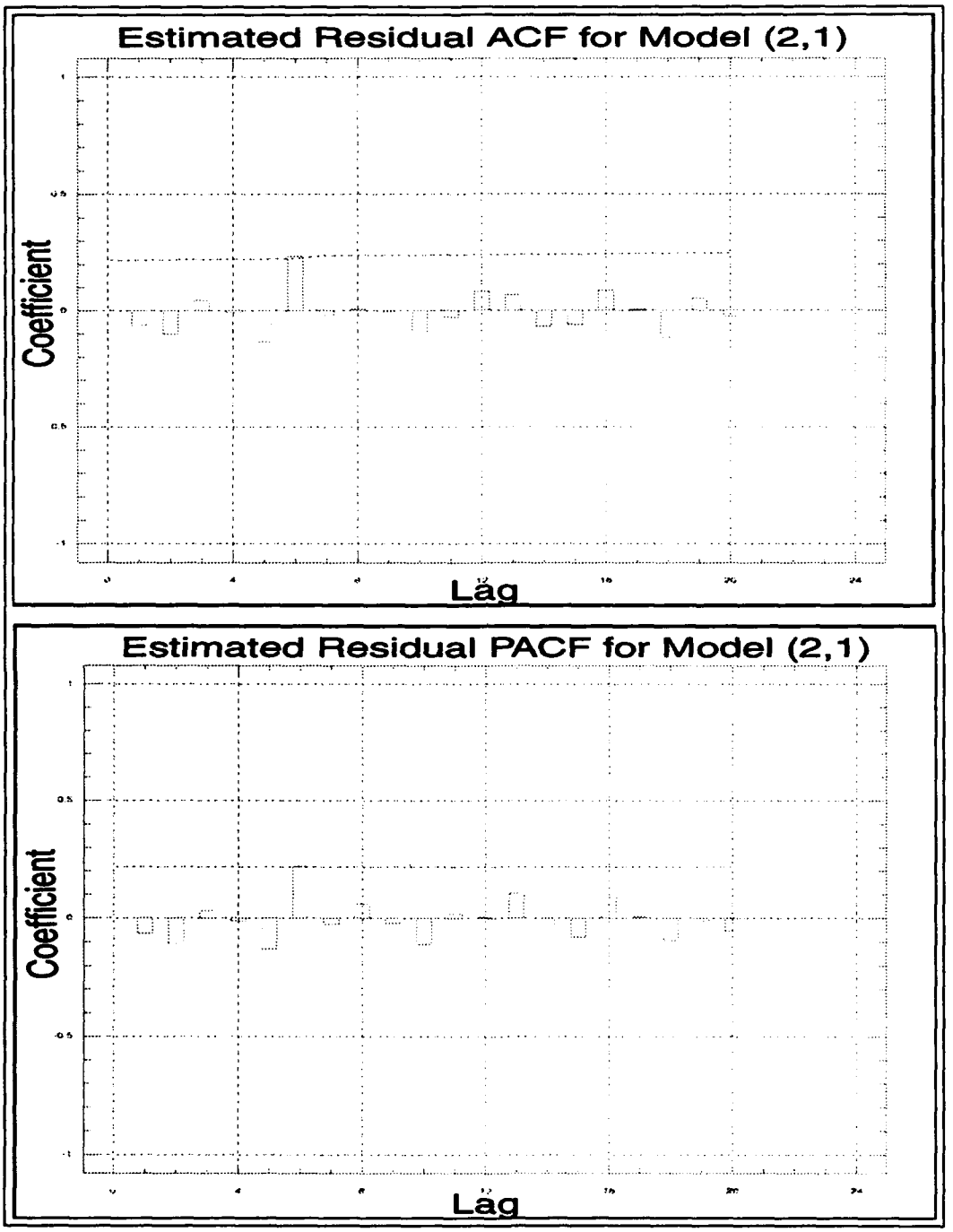

Figure F.63. Plots of the residual ACF and PACF for Model KE(2,0,1).

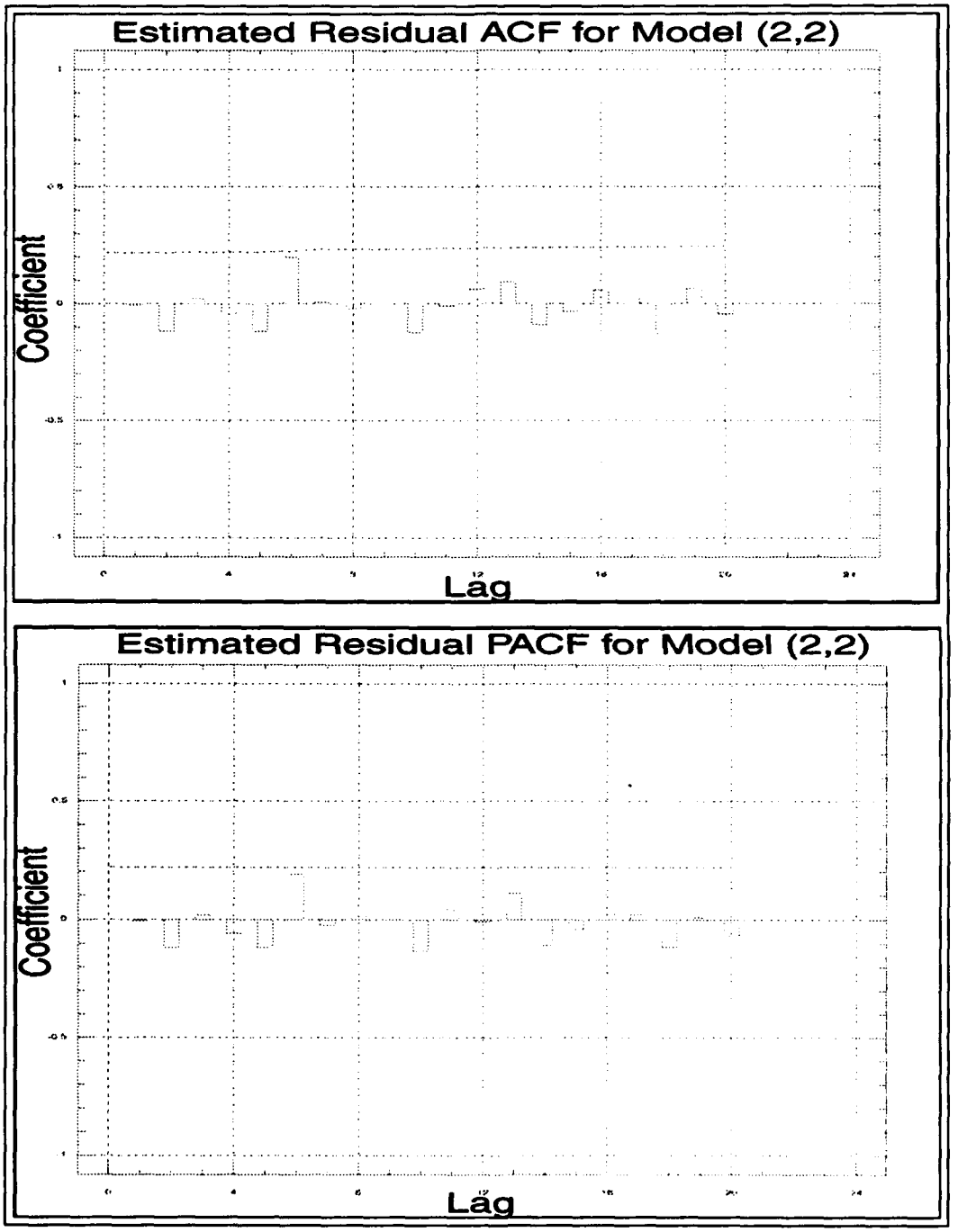

Figure F.64. Plots of the residual ACF and PACF for Model KE(2,0,2).

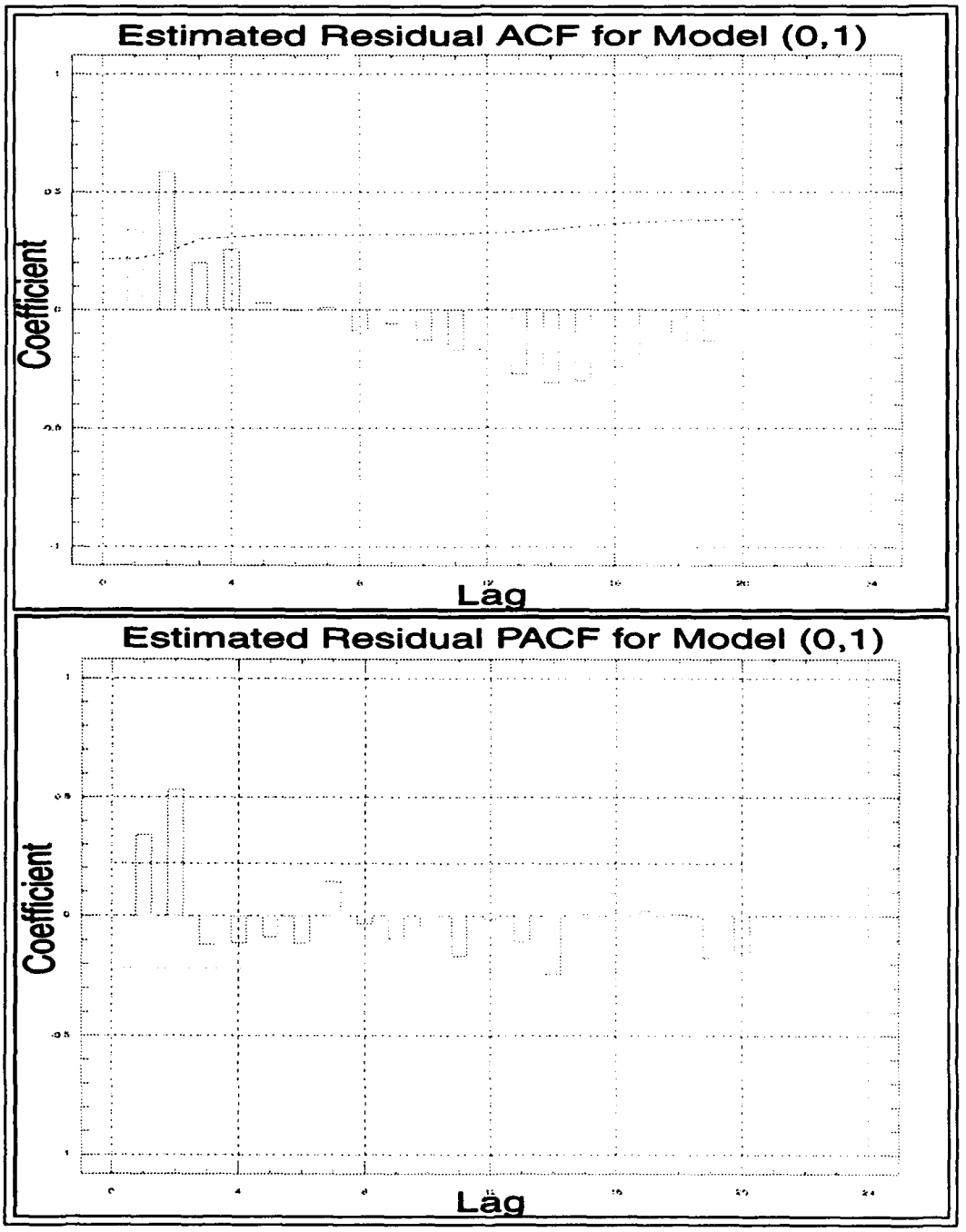

Figure F.65. Plots of the residual ACF and PACF for Model KS(0,0,1).

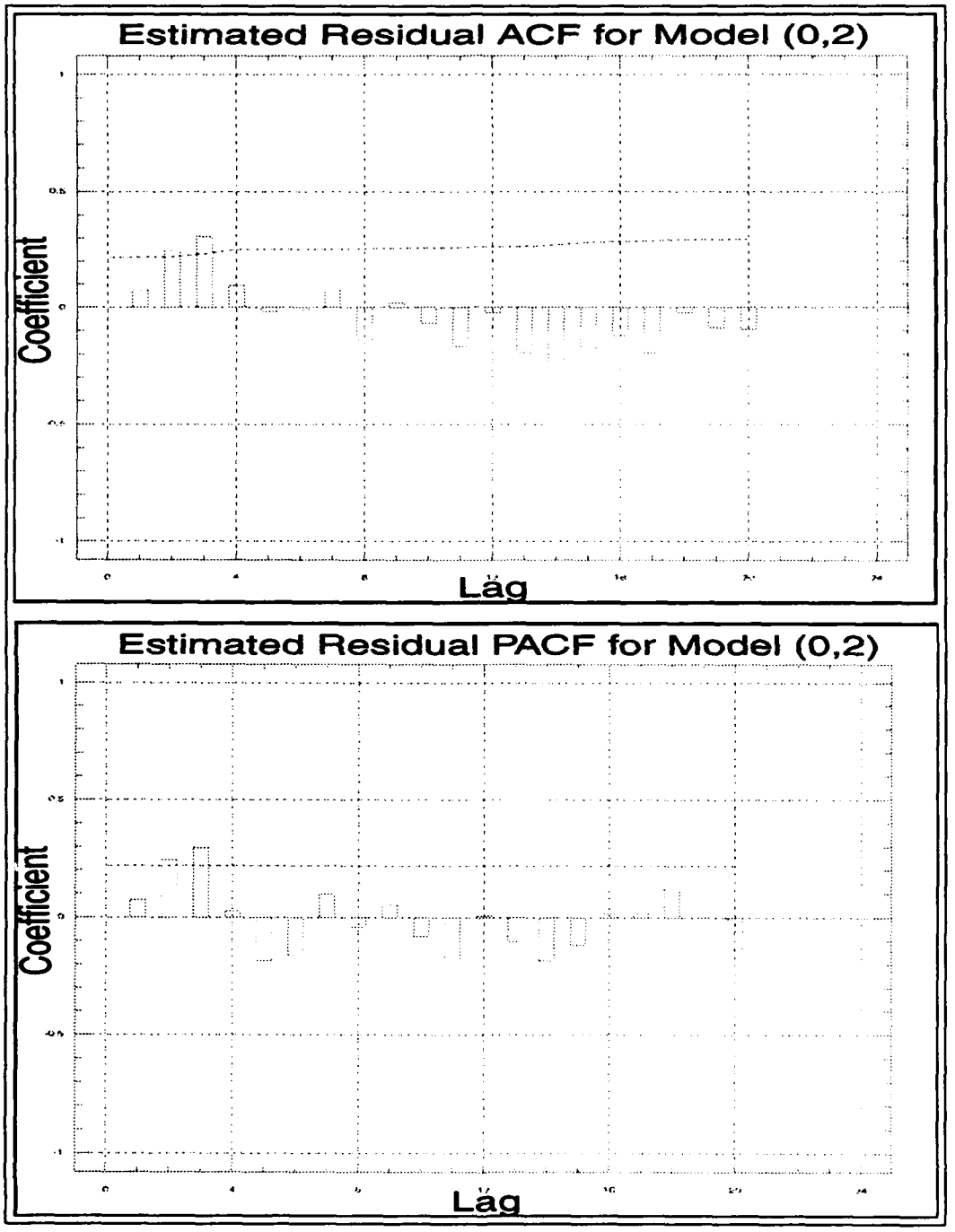

Figure F.66. Plots of the residual ACF and PACF for Model KS(0,0,2).

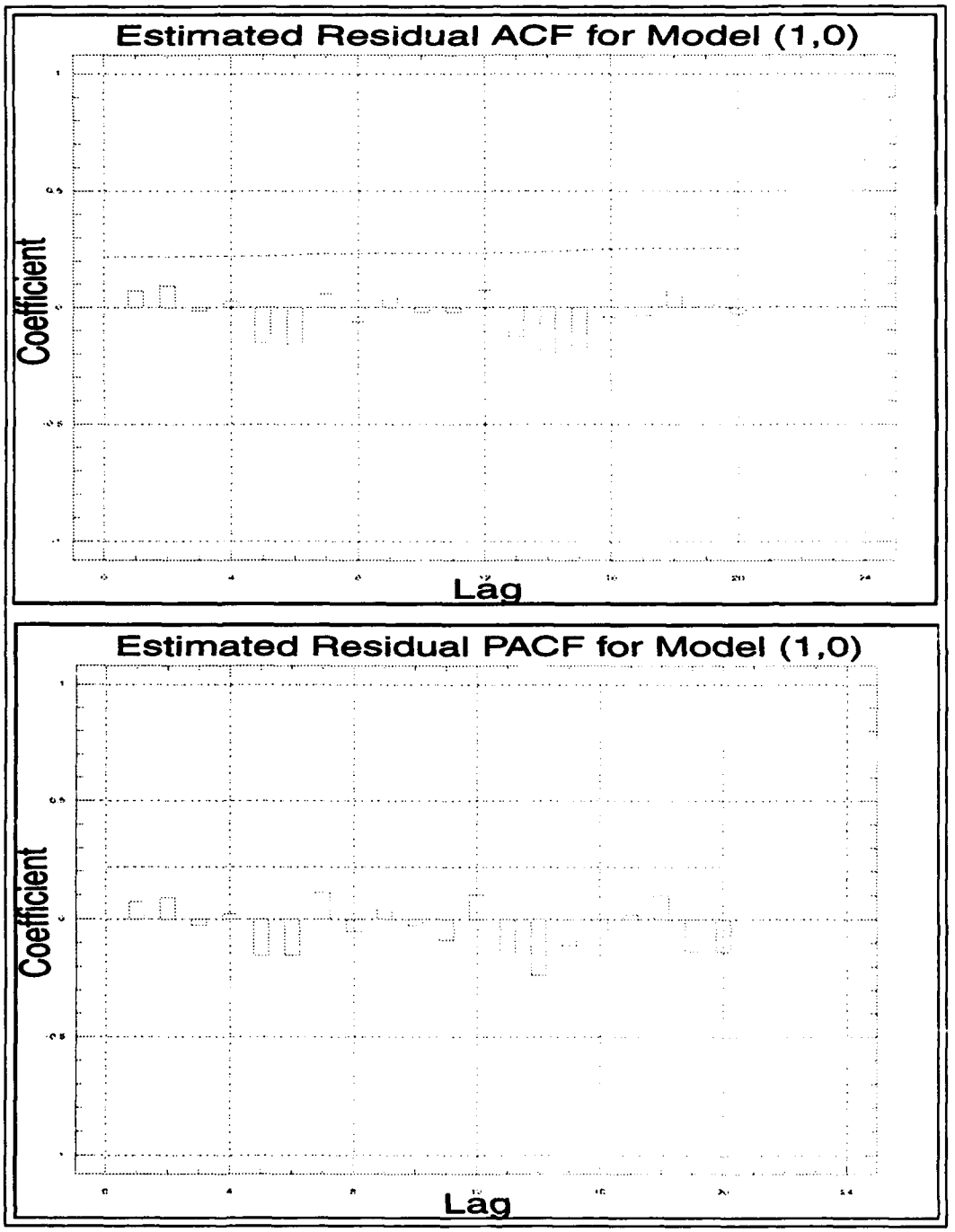

Figure F.67. Plots of the residual ACF and PACF for Model KS(1,0,0).

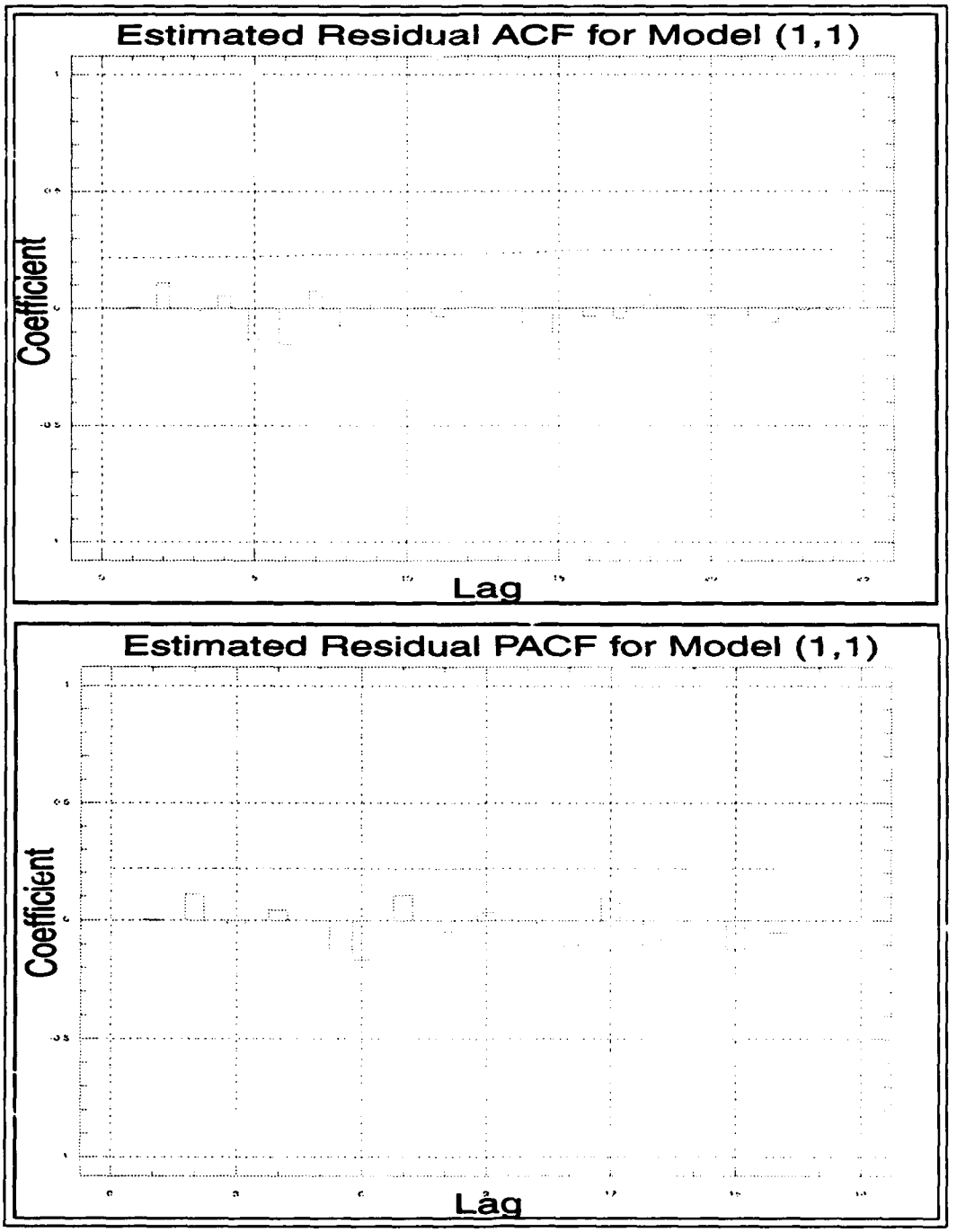

Figure F.68. Plots of the residual ACF and PACF for Model KS(1,0,1).

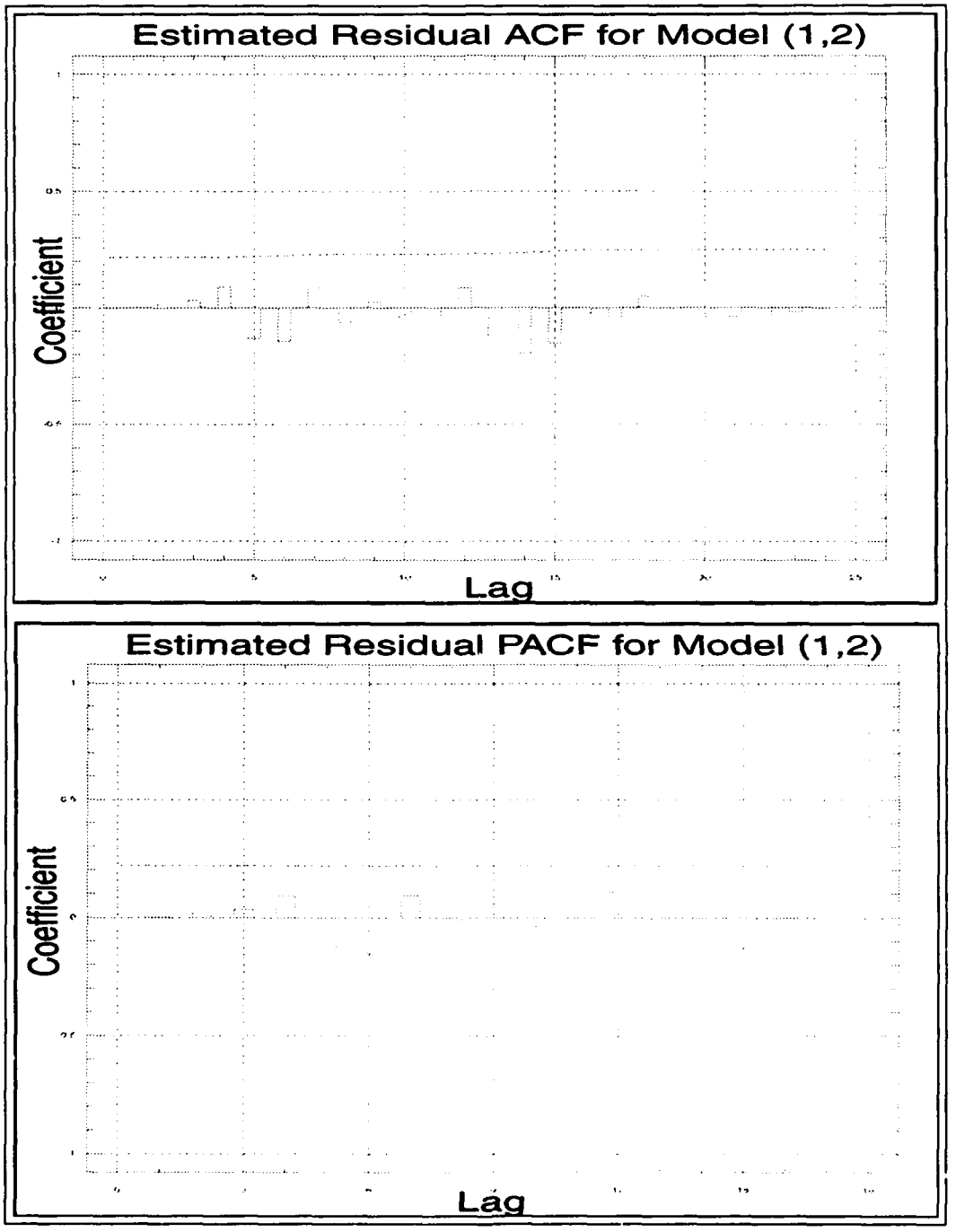

Figure F.69. Plots of the residual ACF and PACF for Model KS(1,0,2).

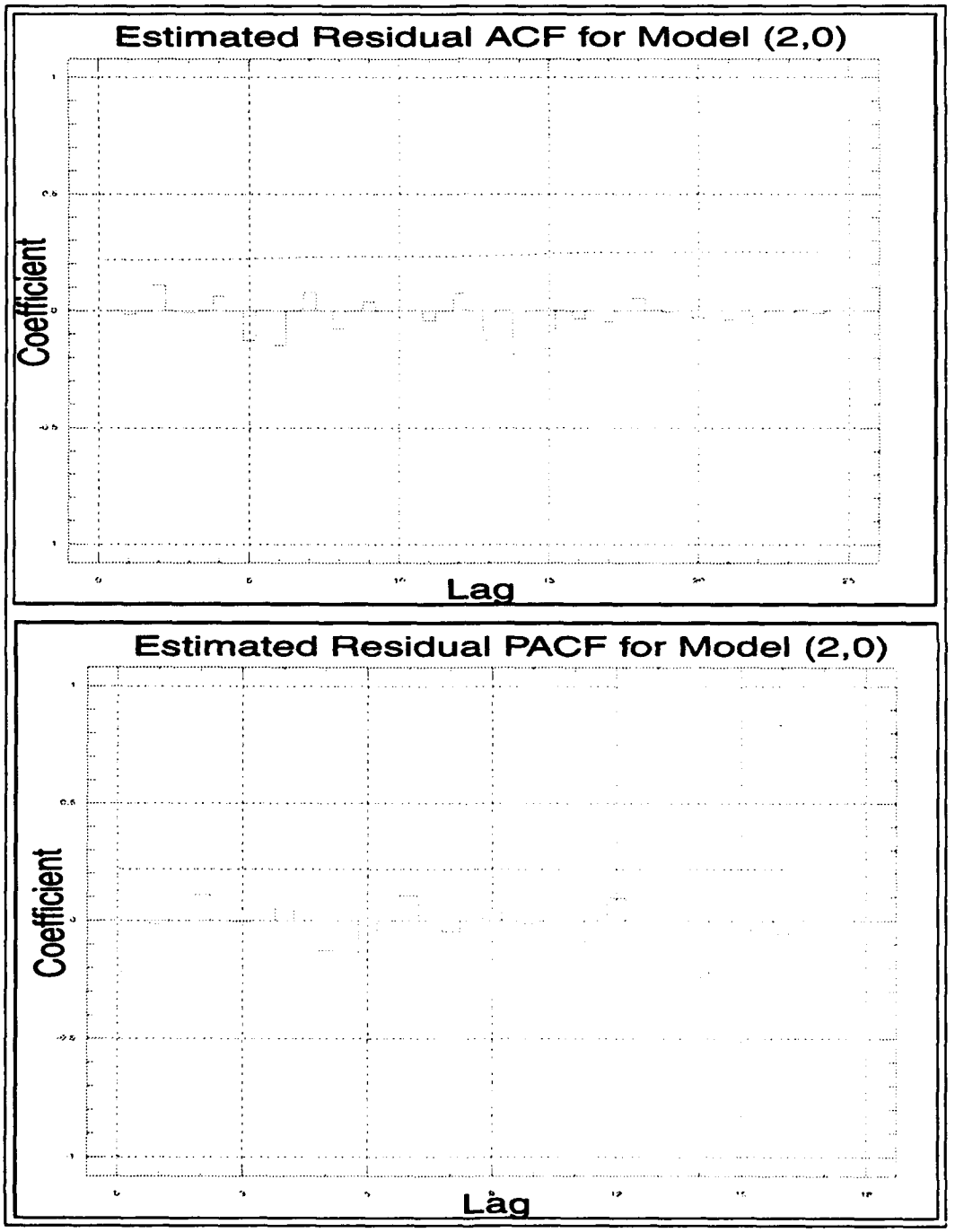

Figure F.70. Plots of the residual ACF and PACF for Model KS(2,0,0).

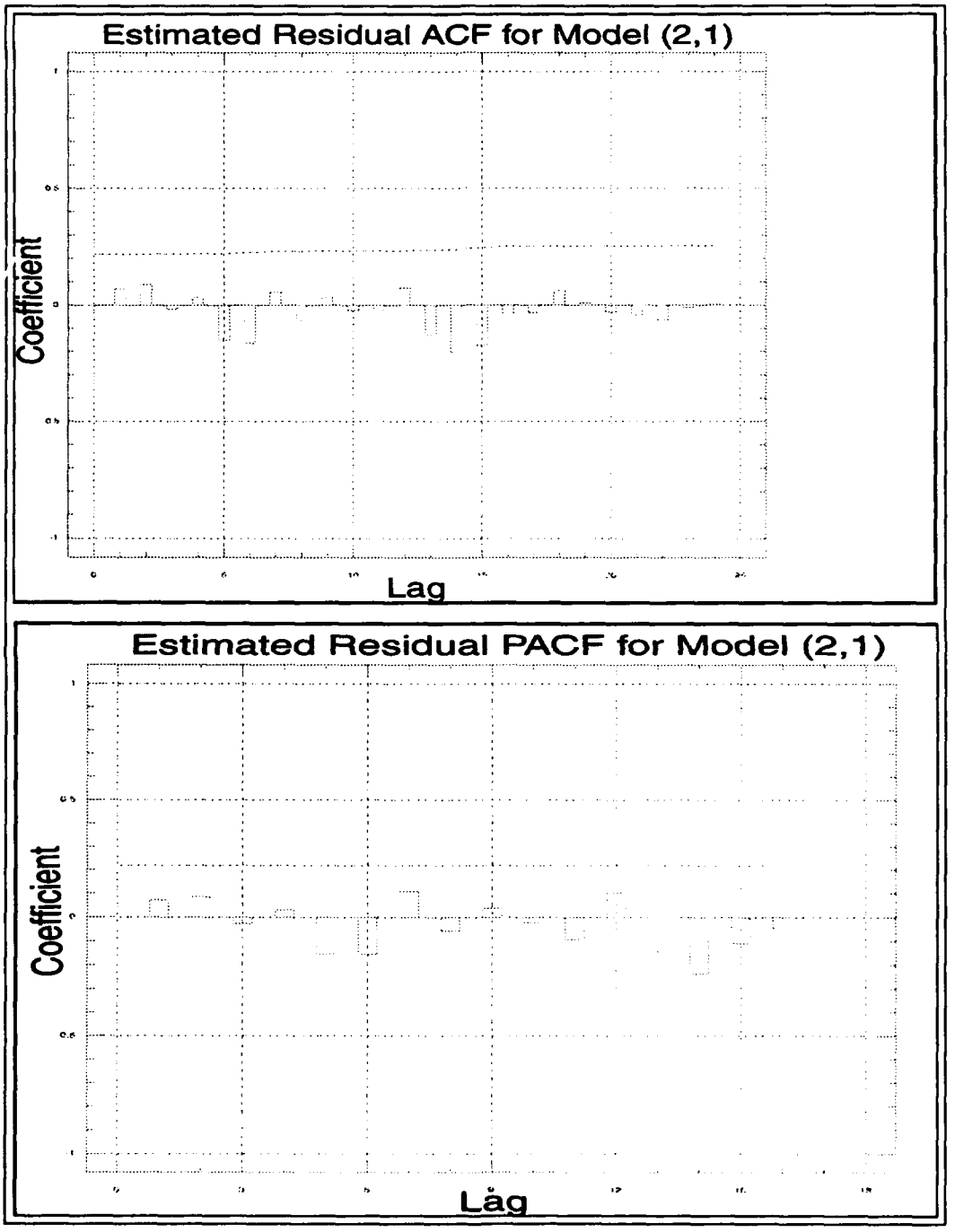

Figure F.71. Plots of the residual ACF and PACF for Model KS(2,0,1).

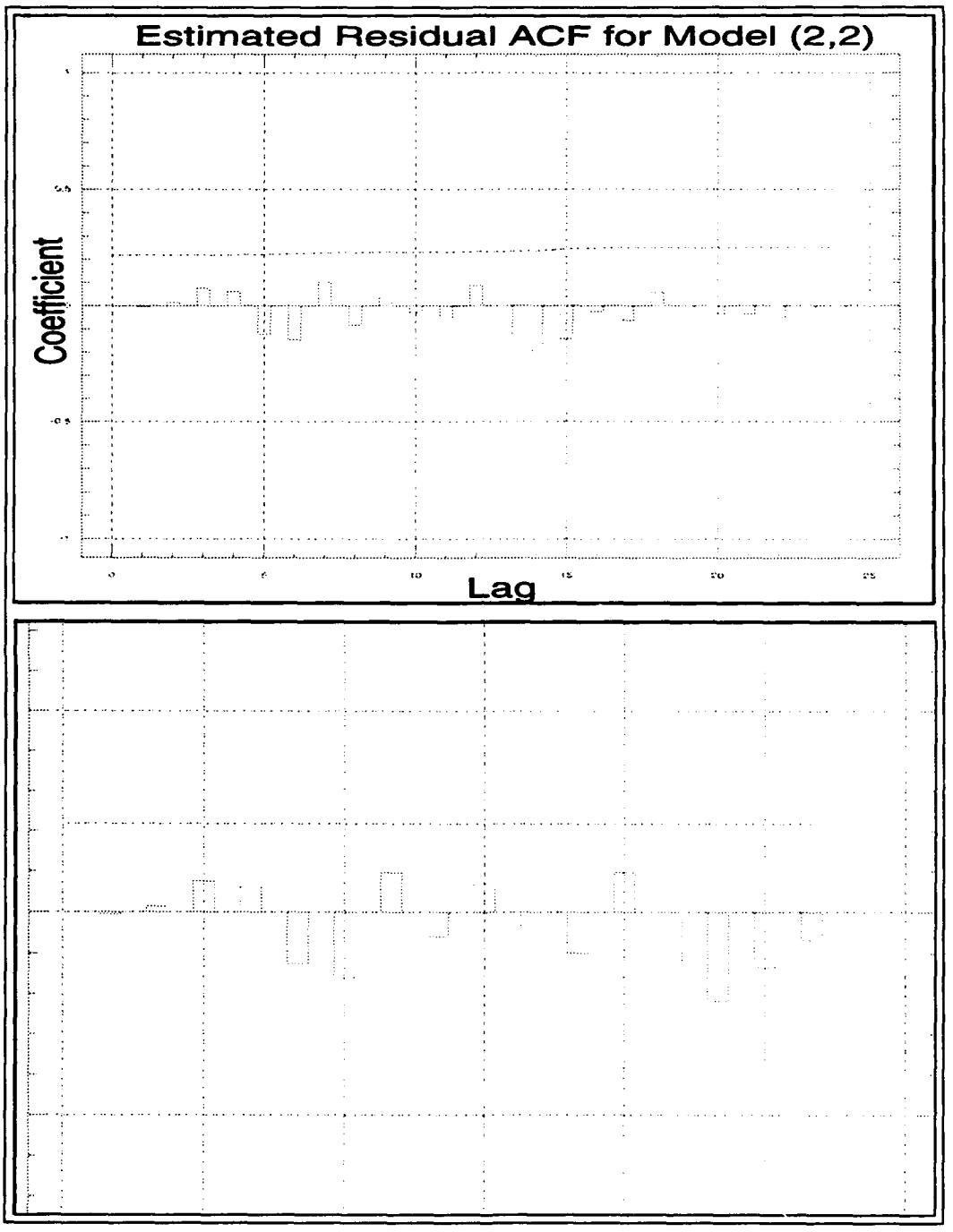

Figure F.72. Plots of the residual ACF and PACF for Model KS(2,0,2).

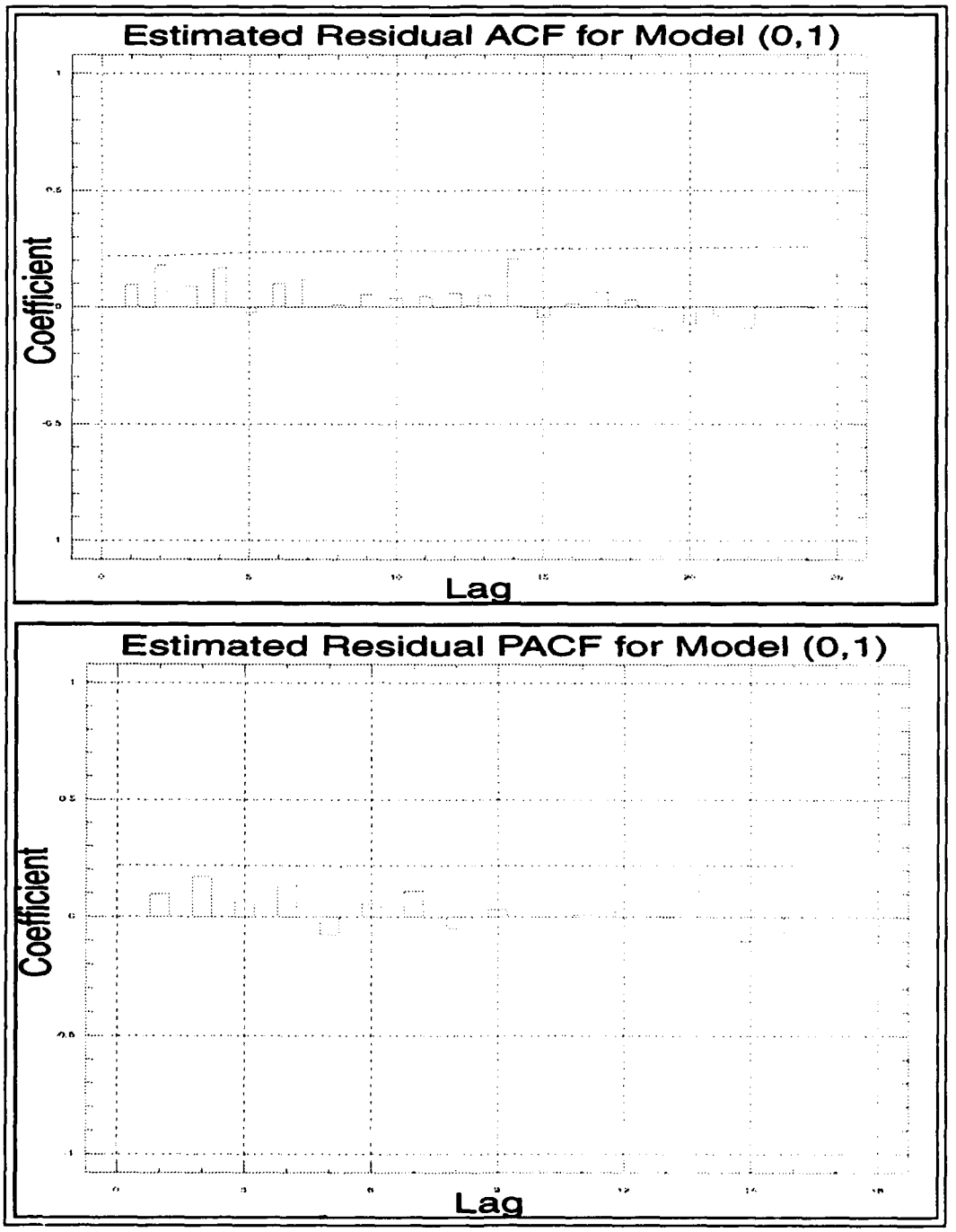

Figure F.73. Plots of the residual ACF and PACF for Model KW(0,0,1).

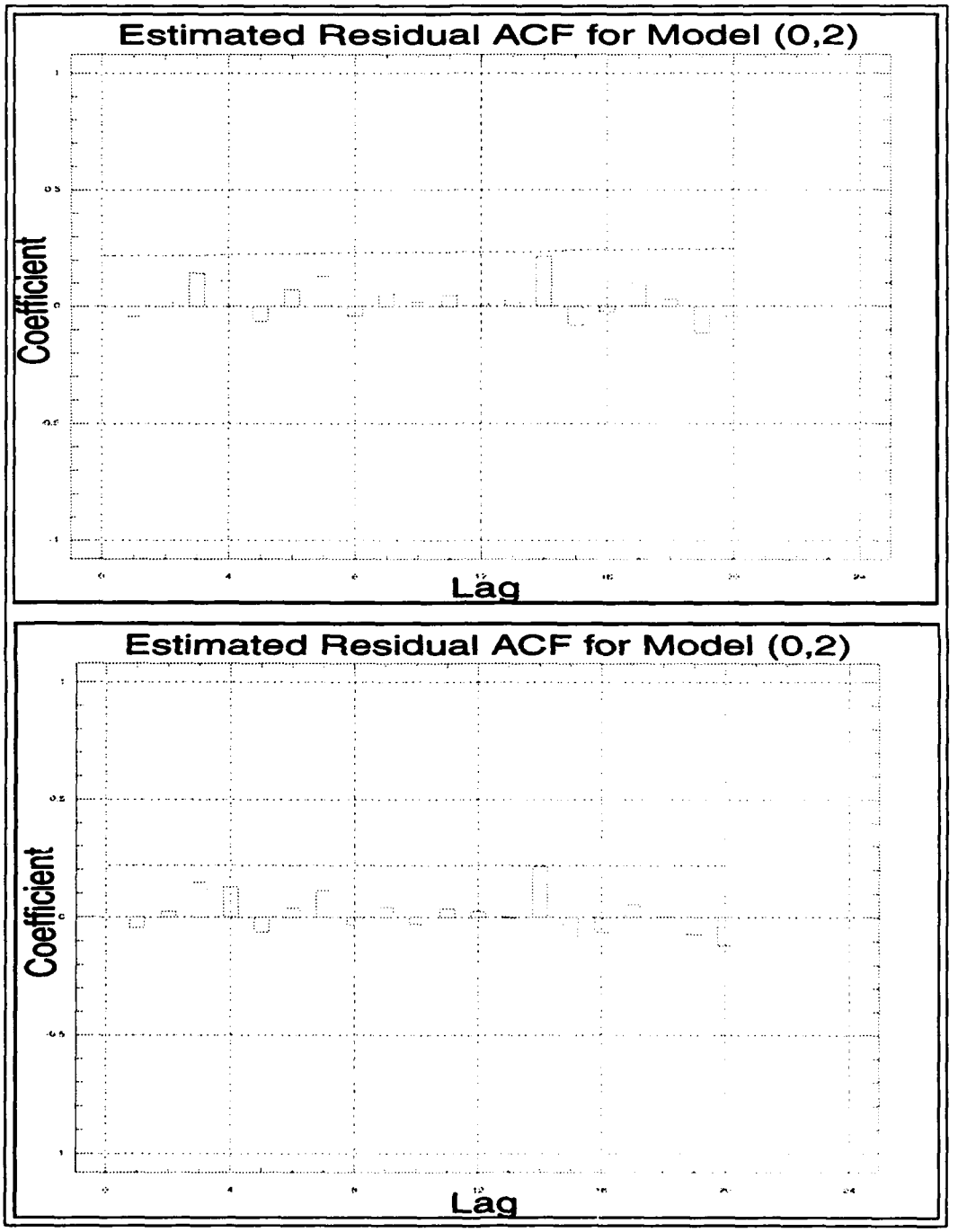

Figure F.74. Plots of the residual ACF and PACF for Model KW(0,0,2).

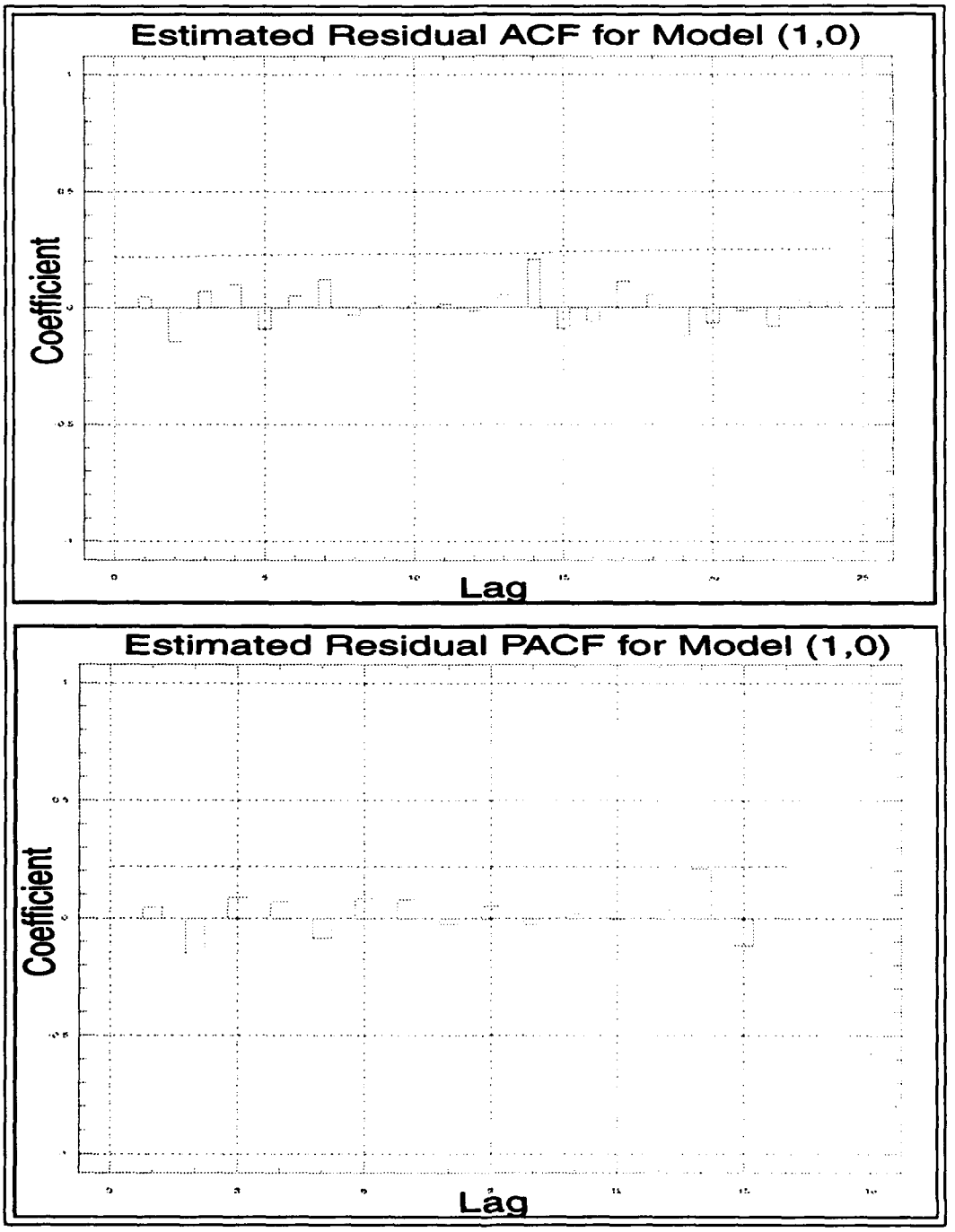

Figure F.75. Plots of the residual ACF and PACF for Model KW(1,0,0).

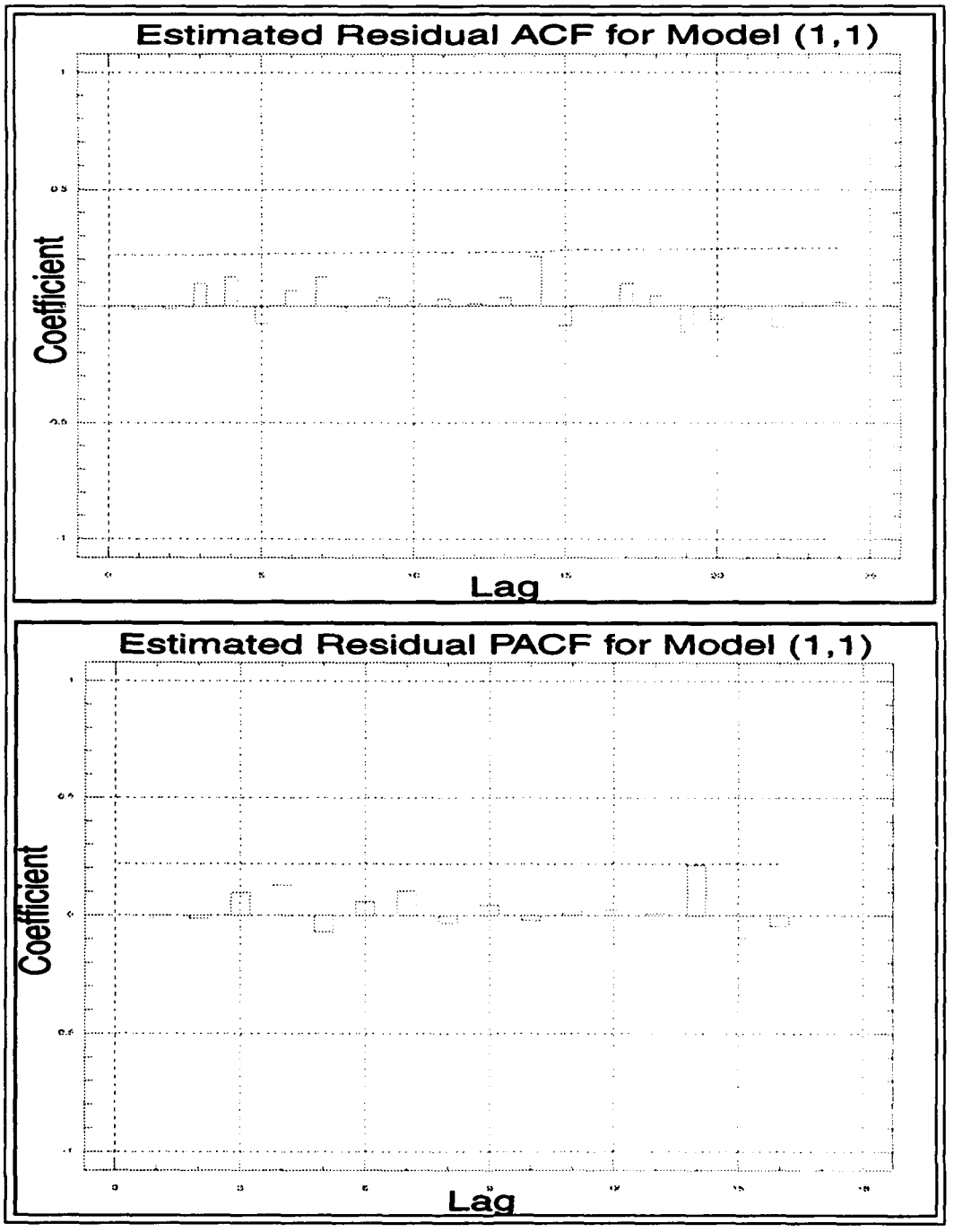

Figure F.76. Plots of the residual ACF and PACF for Model KW(1,0,1).

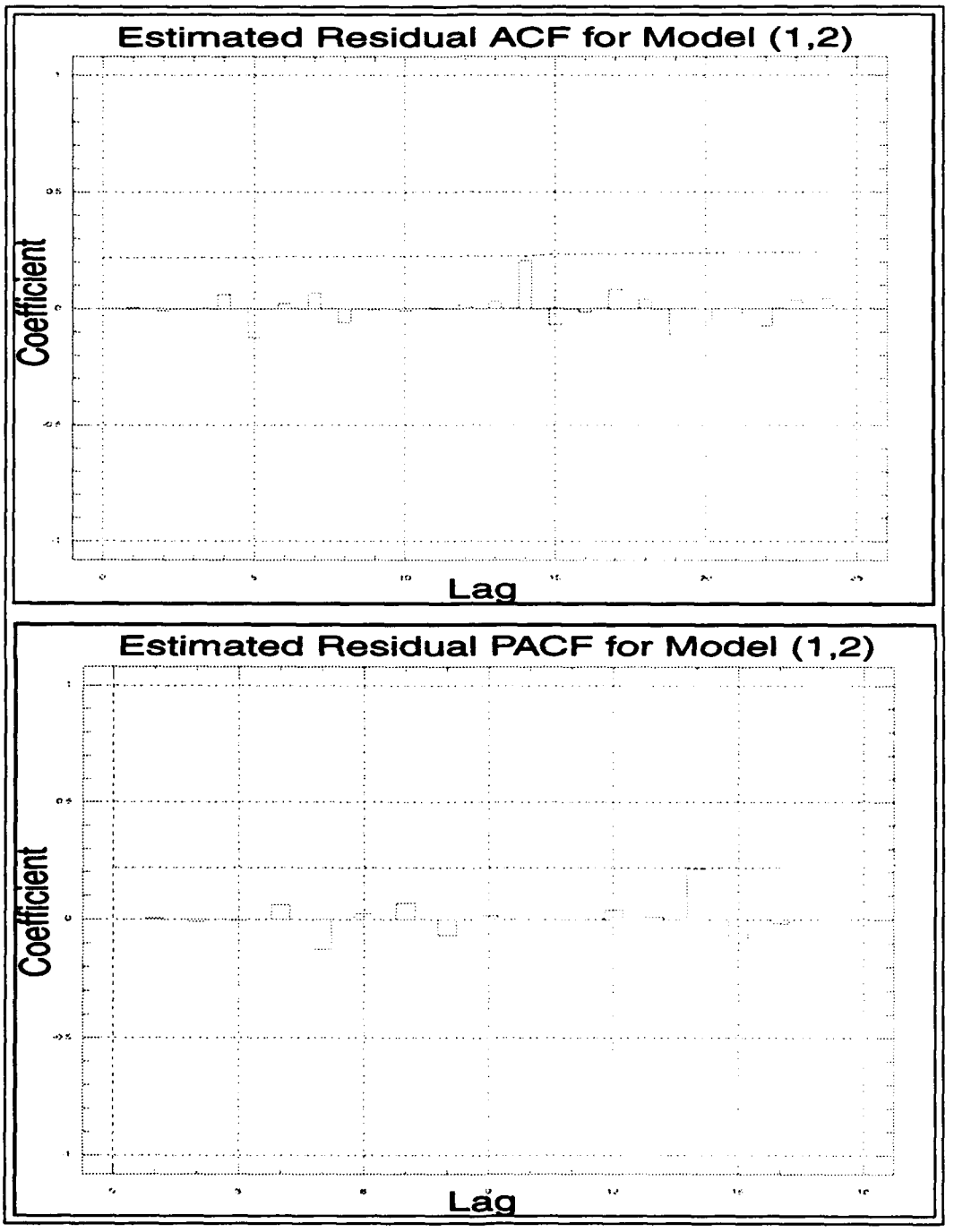

Figure F.77. Plots of the residual ACF and PACF for Model KW(1,0.2).

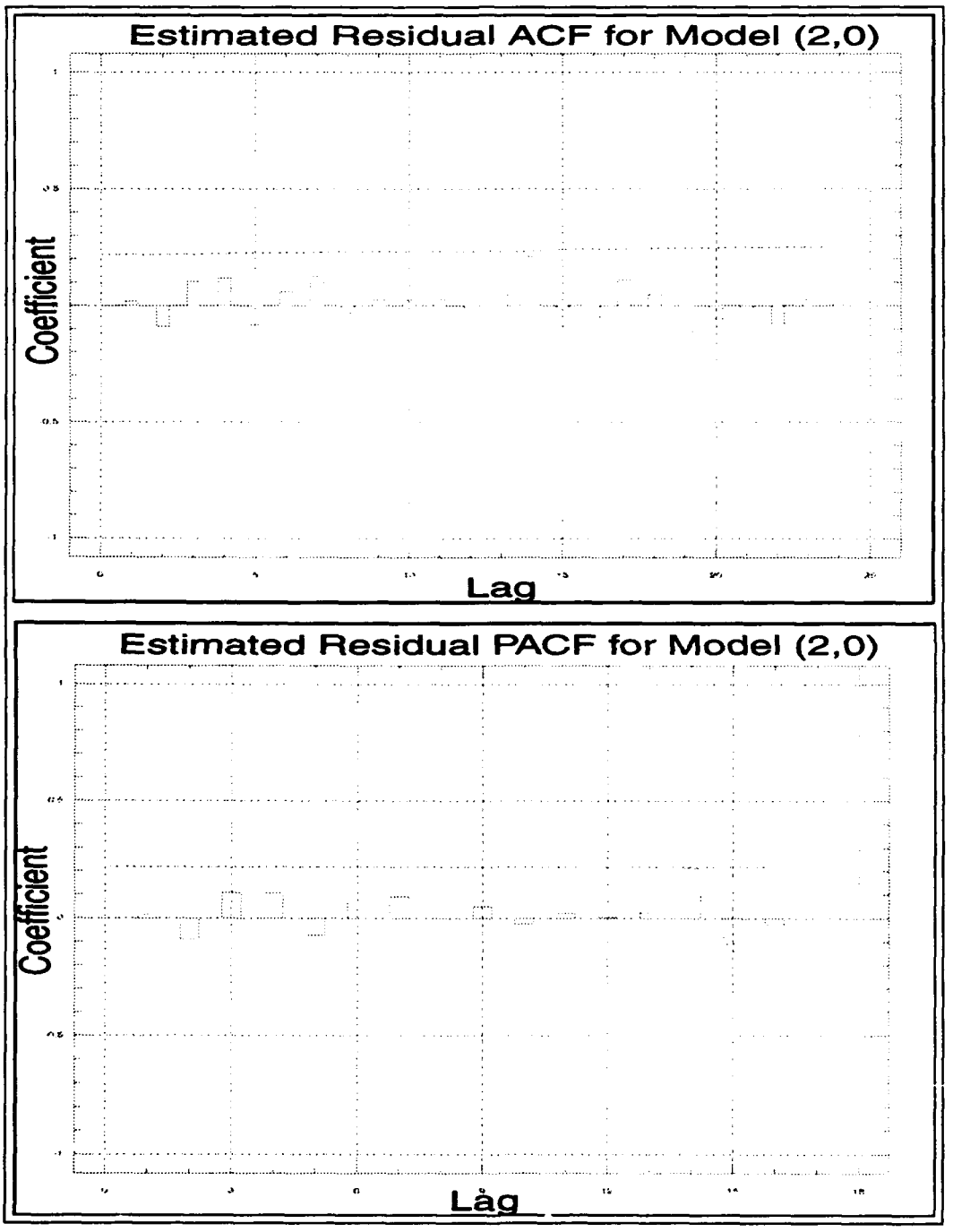

Figure F.78. Plots of the residual ACF and PACF for Model KW(2,0,0).

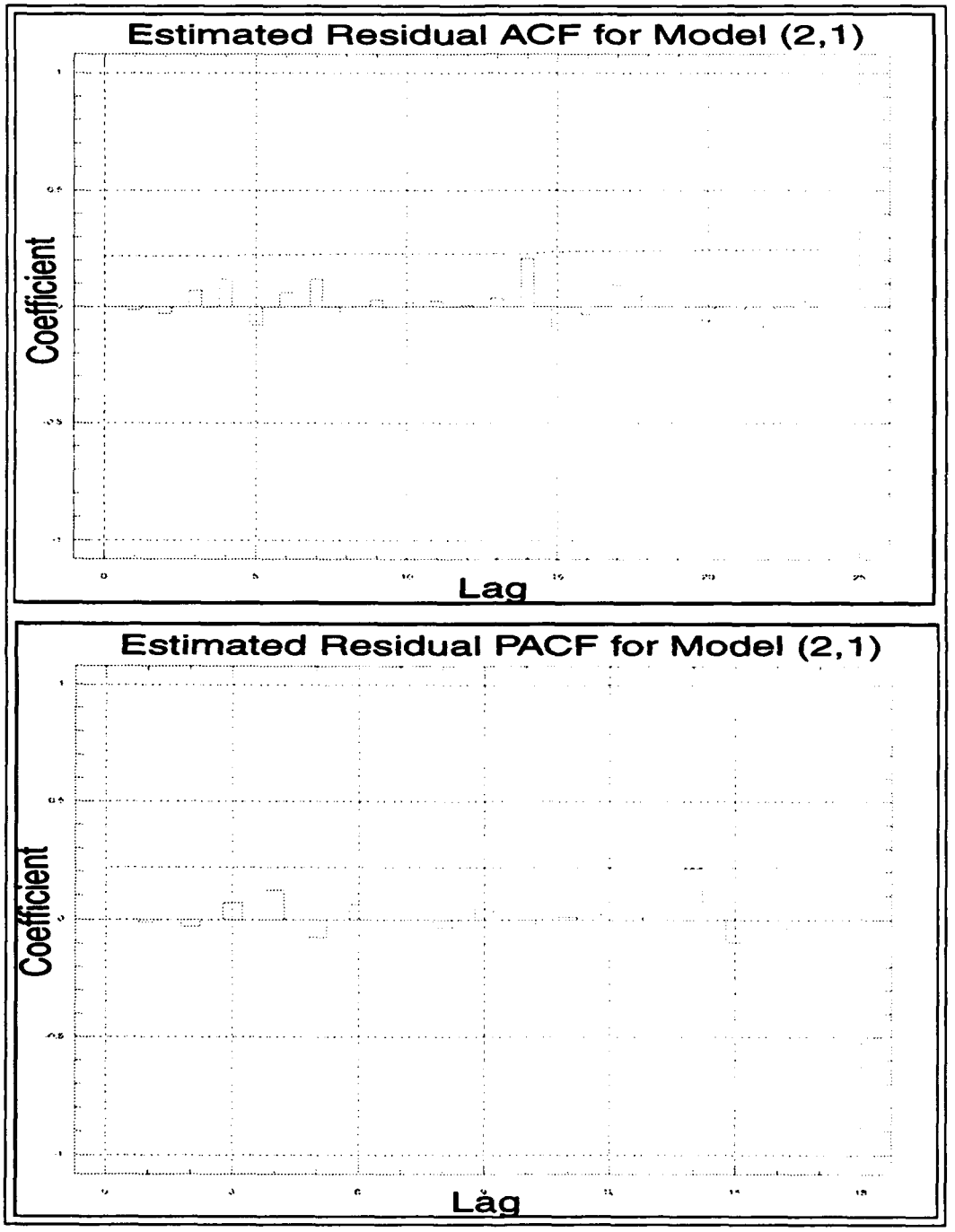

Figure F.79. Plots of the residual ACF and PACF for Model KW(2,0,1).

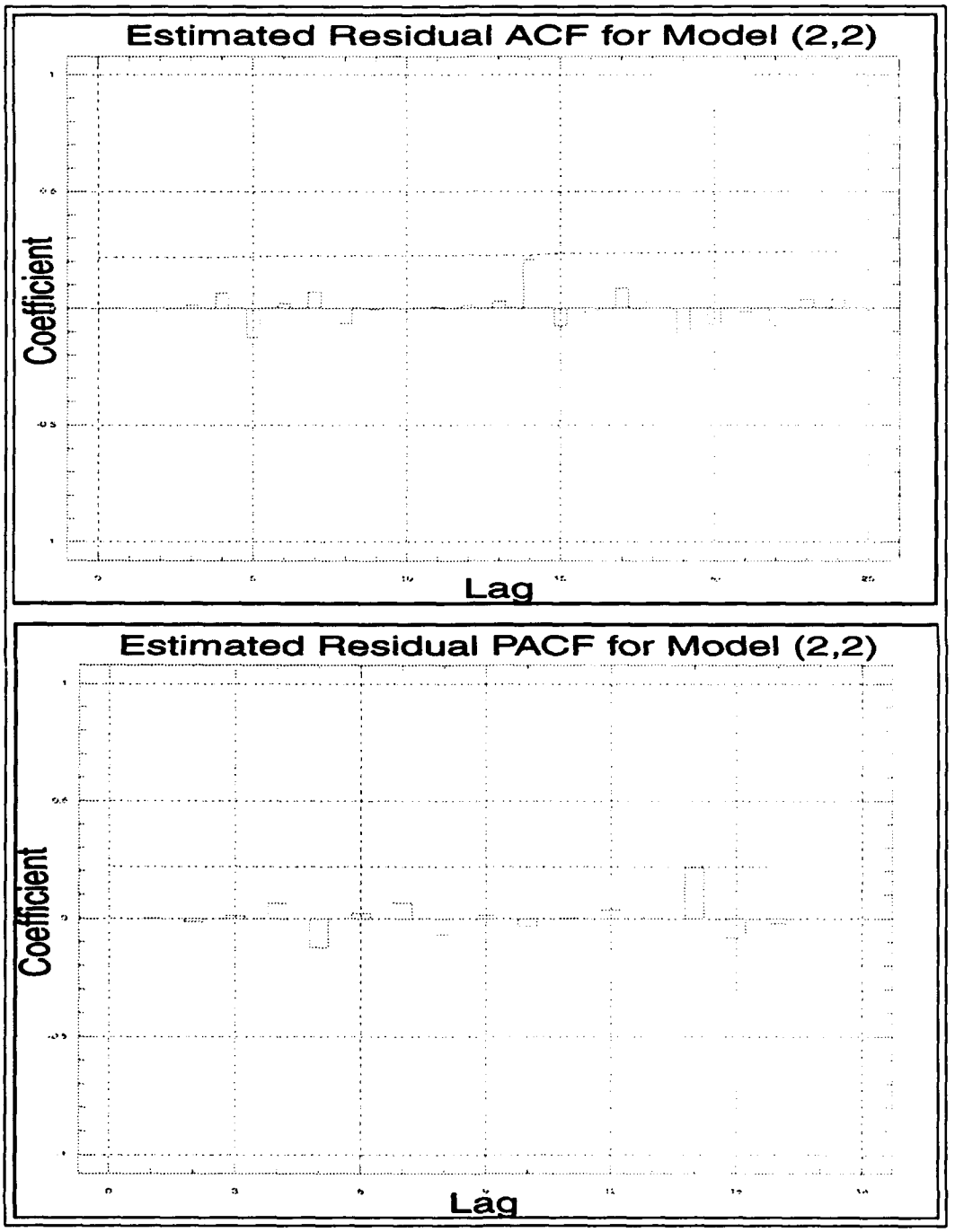

Figure F.80. Plots of the residual ACF and PACF for Model KW(2,0,2).

Cumulative Probability Plots and Periodograms of Residuals (by model)

Cumulative Probability Plots. The cumulative probability plot[3] provides a visual means to test for randomness (white noise). The cumulative probability plot is particularly useful because it is sensitive to periodic effects for which ACFs are not sensitive. The Statgraphics INTPER routine plots the cumulative sum of the periodogram ordinates normalized to a (0,1) vertical scale. The plot includes 75 percent and 95 percent Kolmogorov-Smirnov bounds for a uniform distribution of ordinates as an approximate test for model inadequacy (27:I-5). If the model is adequate, the normalized cumulative periodogram, $C(f_i)$, plotted against the frequency, f_i, will be scattered points about a straight line joining the points (0,0) and (0.5,1). Conversely, inadequate models result in plots that cross the Kolmogorov-Smirnov bounds (3:295). The value of the ordinate, $C(f_i)$ at frequency f_i is given by the formulas.[4]

$$C(f_i) = \frac{n}{2}(a_i^2 + b_i^2)$$

where,

$i = 1, 2, ..., (n-1)/2$

$a_i = \frac{2}{n} \sum_{t=1}^{n} y_t cos(2\pi f_i t)$

$b_i = \frac{2}{n} \sum_{t=1}^{n} y_t (2\pi f_i t)$

$f_i = i/n$

If n is even, an additional term is added:

$$I(0.5) = \frac{1}{n \sum_{t=1}^{n} (-1)^t Y_t^2}$$

[3] Also referred to as a normal probability plot or integrated periodogram.

[4] The formulas presented here are extracted from the Statgraphics Reference Manual (27:I-11). The formulas are simplifications based on work by Box and Jenkins (3:36).

Periodograms. Periodograms are used to test for nonrandom periodic effects in time series. The Statgraphics PER routine uses fast Fourier transforms to calculate the average squared amplitude of the sinusoids for various frequencies. The amplitudes are plotted vertically and the frequencies are plotted on the horizontal axis. The routine scales the periodogram so that if the mean of the series is 0, the sum of the periodogram ordinates equals the sum of the squared data values. Like cumulative probability plots, periodograms are particularly useful because they are sensitive to periodic effects for which ACFs are not sensitive (3:P-49). The value of the ordinate, $I(f_i)$ at frequency f_i is computed the same as $C(f_i)$ above.

The cumulative probability plots and periodograms are presented by site (Columbia first, then Kirtland), direction (north, east, south, west), and model (numeric order). The site and direction are abbreviated to the first letter in the model name. For example, the Columbia north model (001) is abbreviated as CN(001).

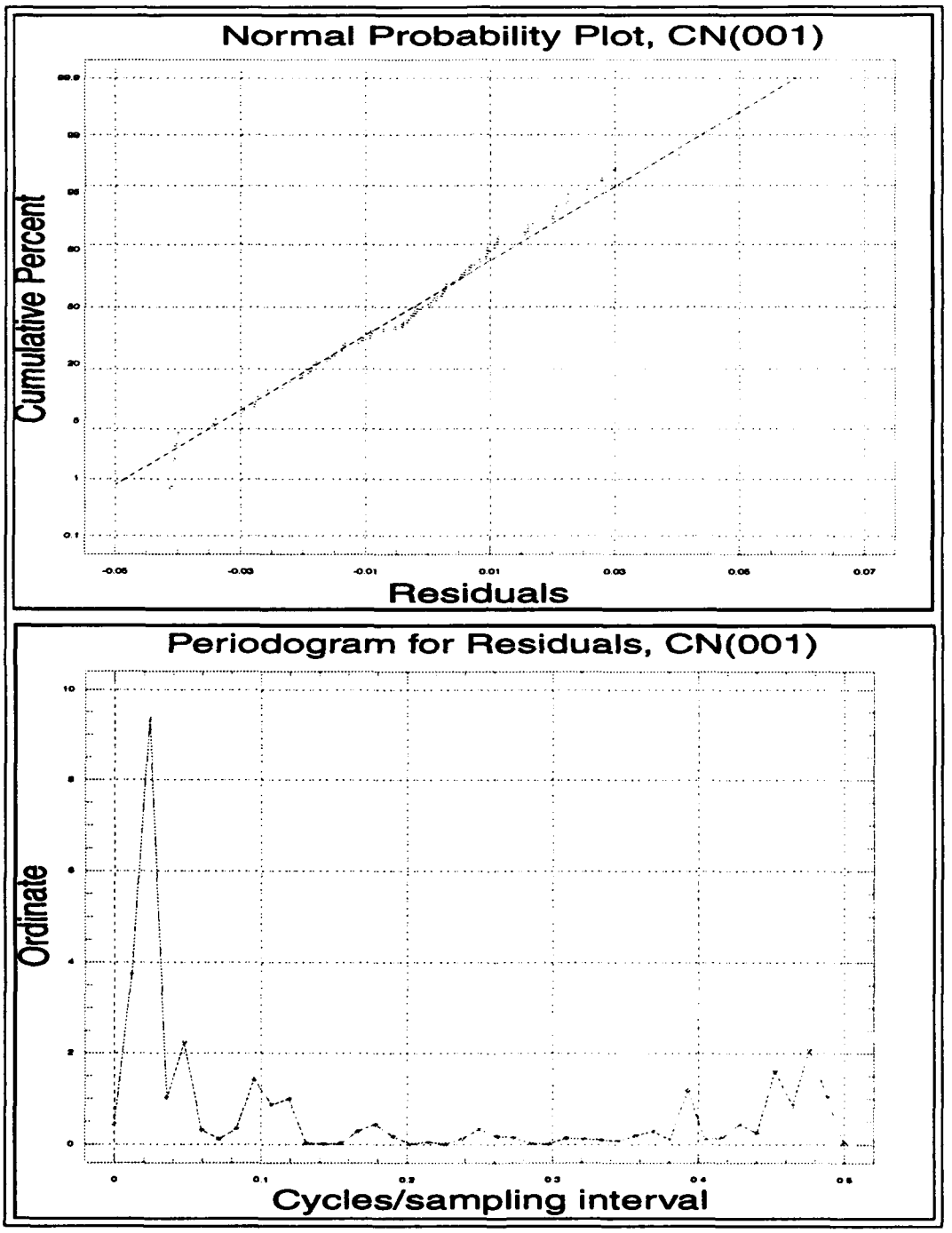

Figure F.81. Cumulative probability plots and periodograms for Model CN(0,0,1).

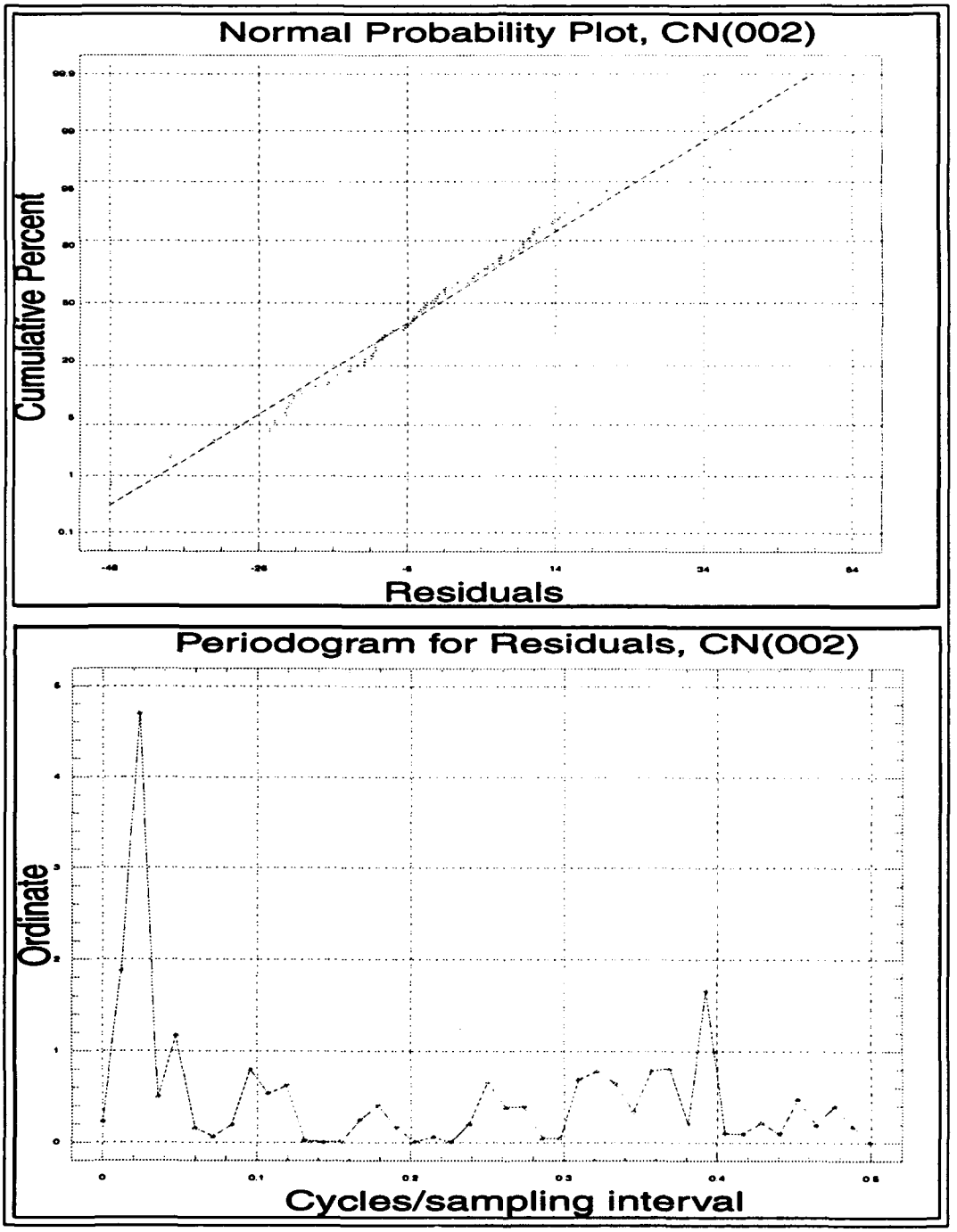

Figure F.82. Cumulative probability plots and periodograms for Model CN(0,0,2).

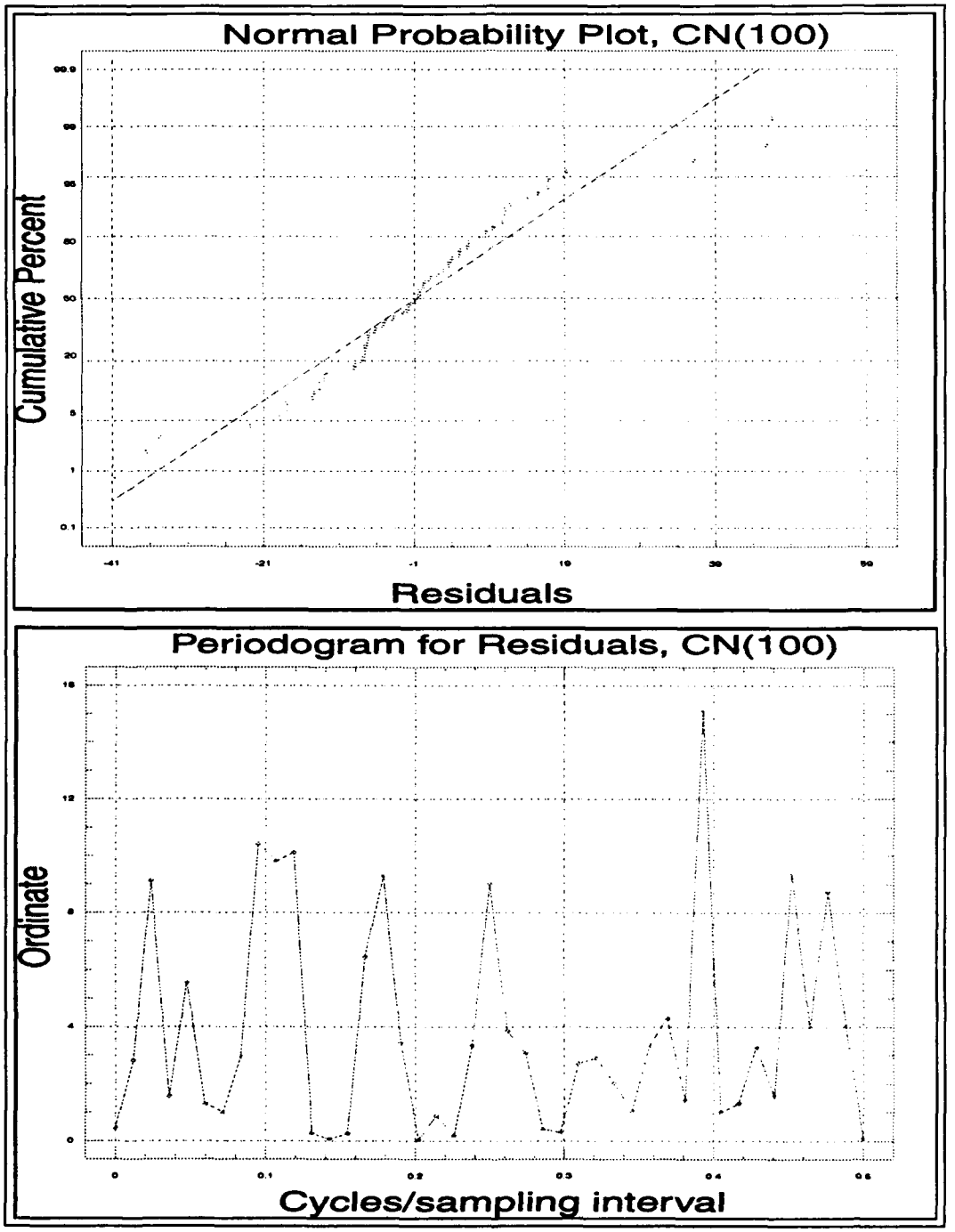

Figure F.83. Cumulative probability plots and periodograms for Model CN(1,0,0).

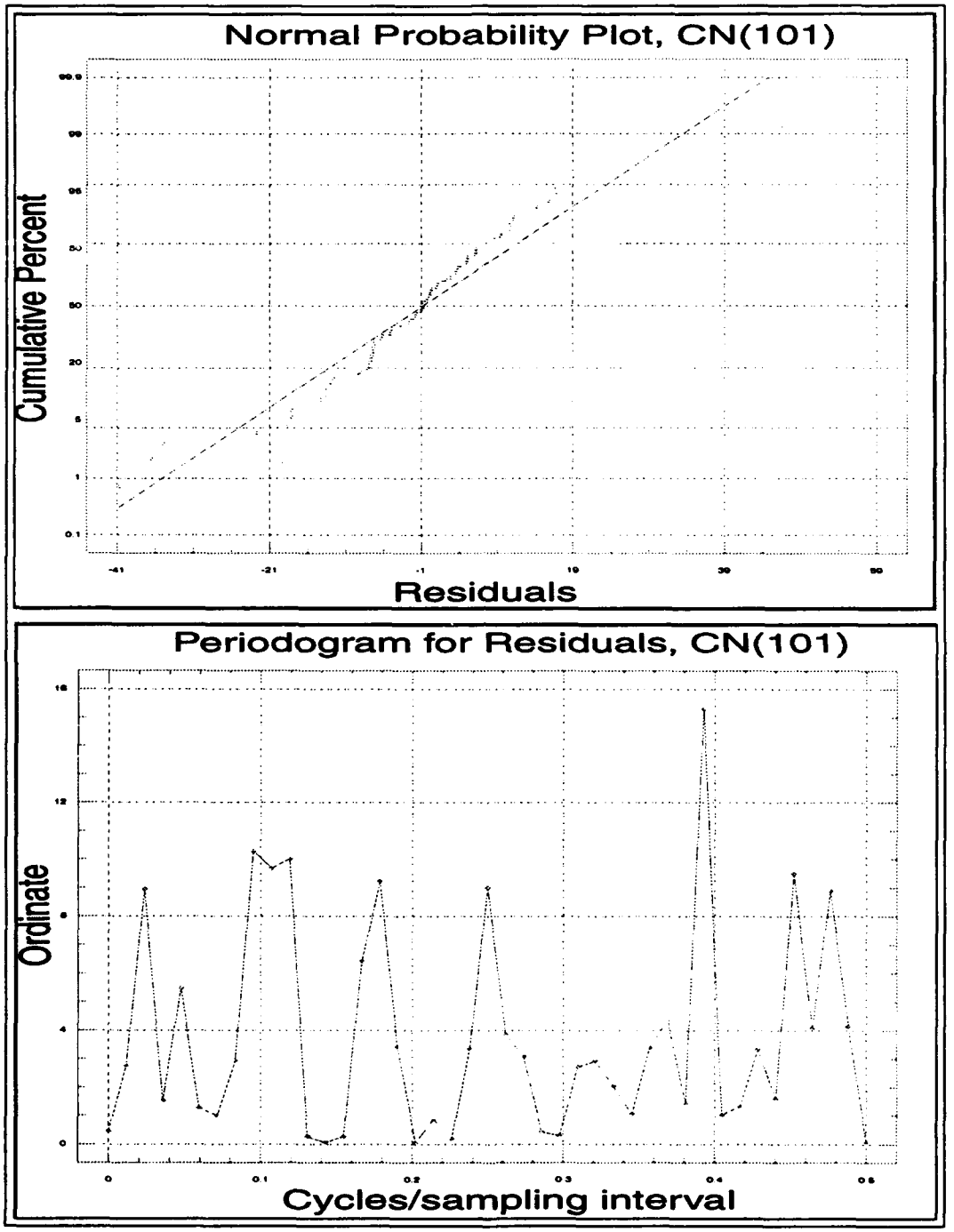

Figure F.84. Cumulative probability plots and periodograms for Model CN(1,0,1).

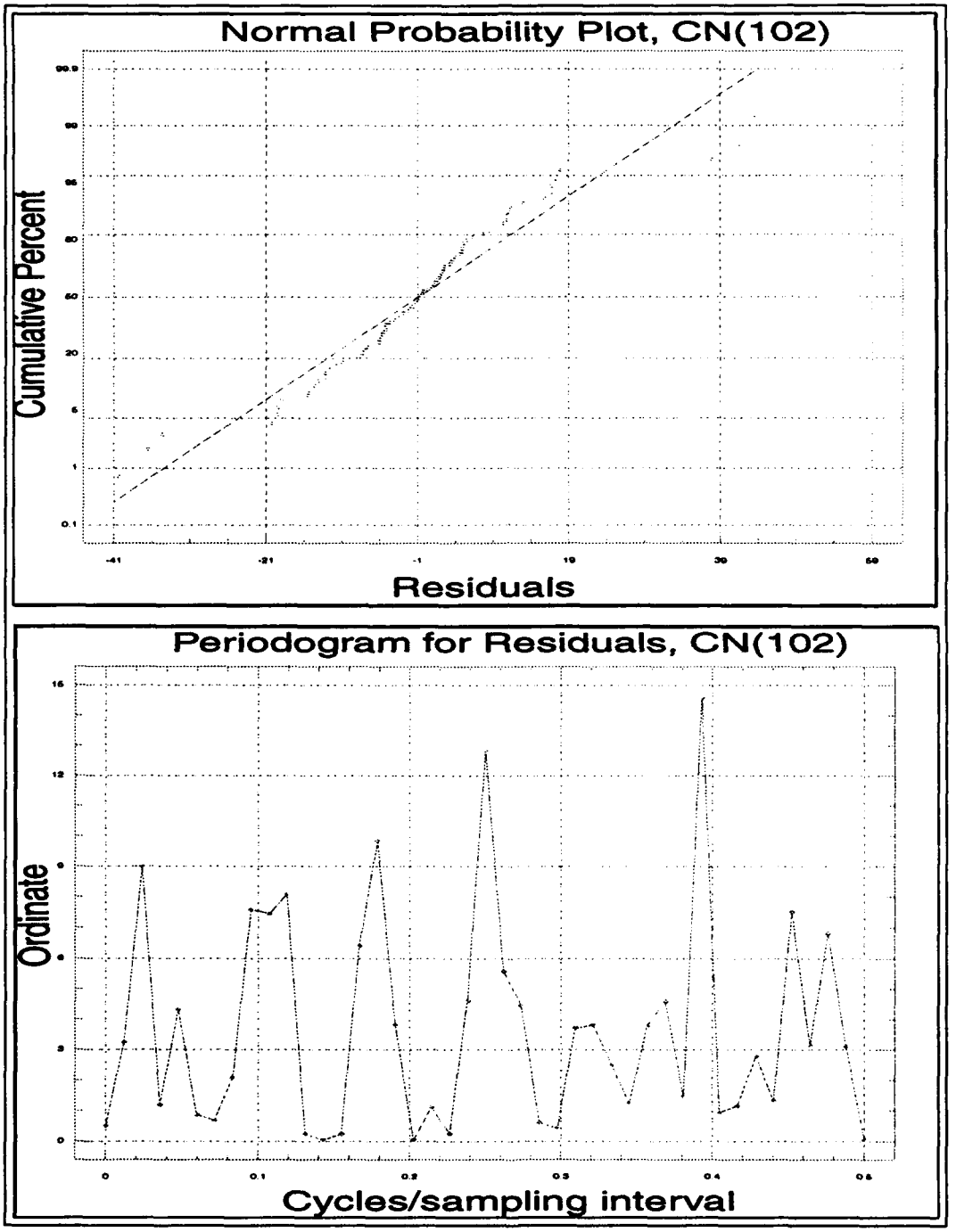

Figure F.85. Cumulative probability plots and periodograms for Model CN(1,0,2).

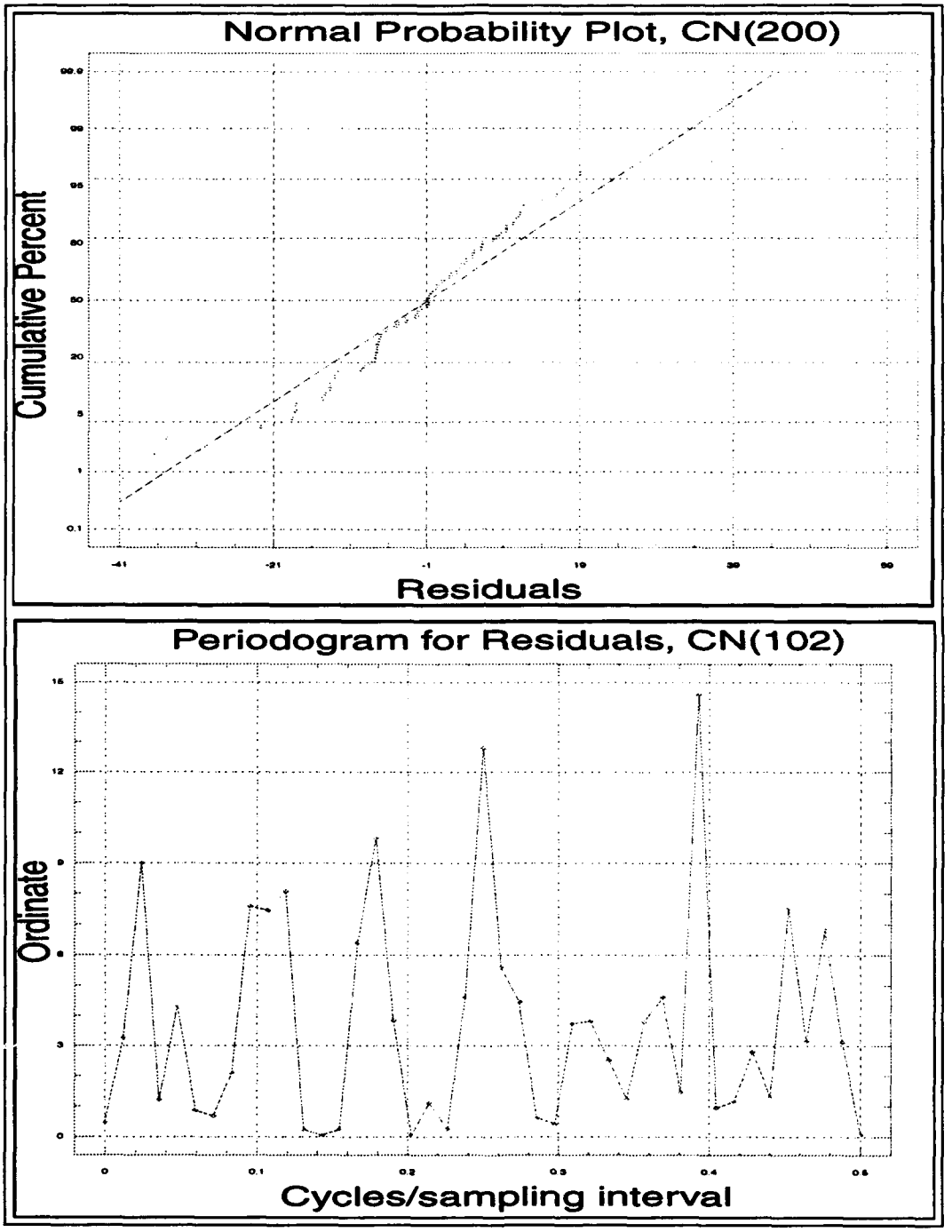

Figure F.86. Cumulative probability plots and periodograms for Model CN(2,0,0).

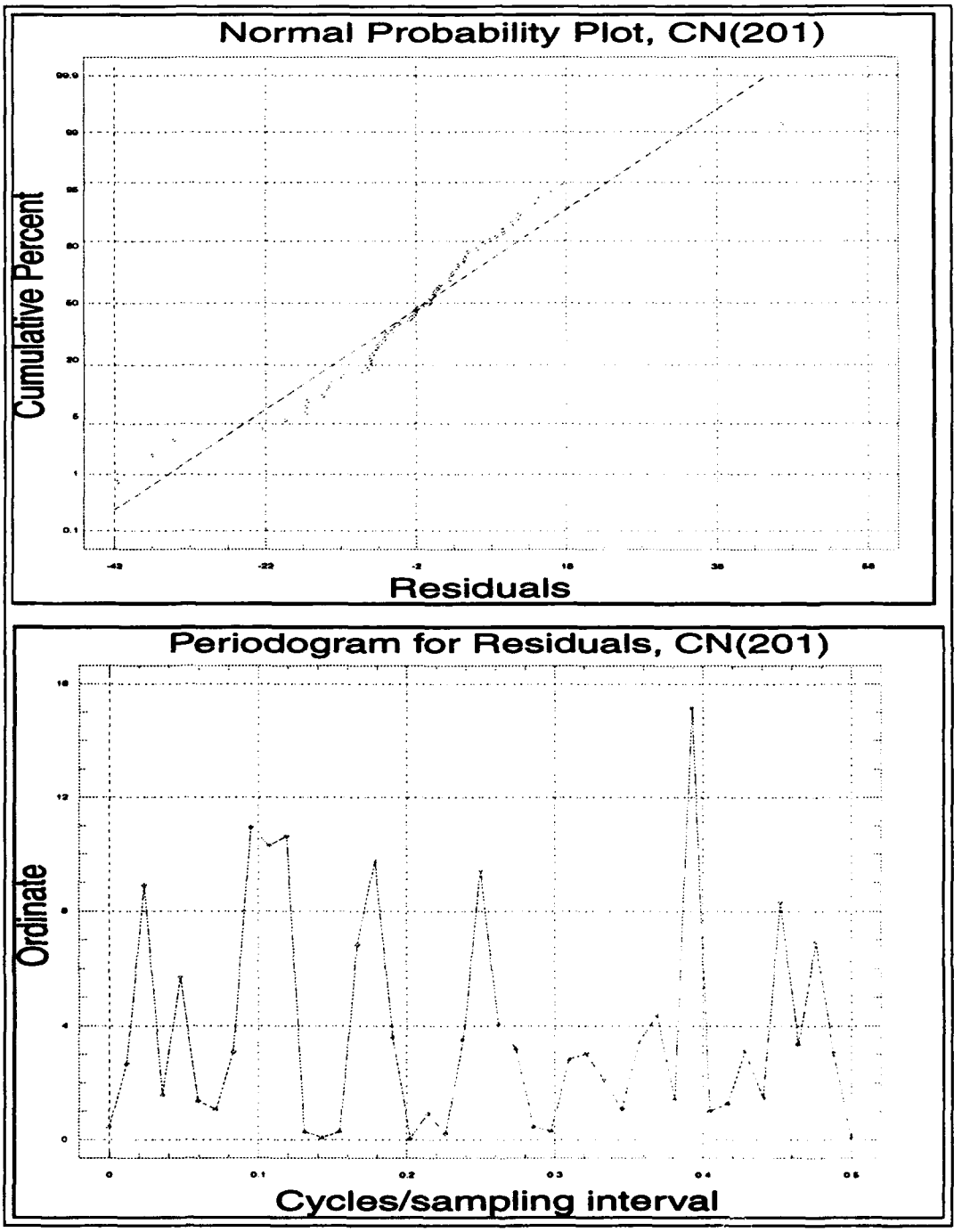

Figure F.87. Cumulative probability plots and periodograms for Model CN(2,0,1).

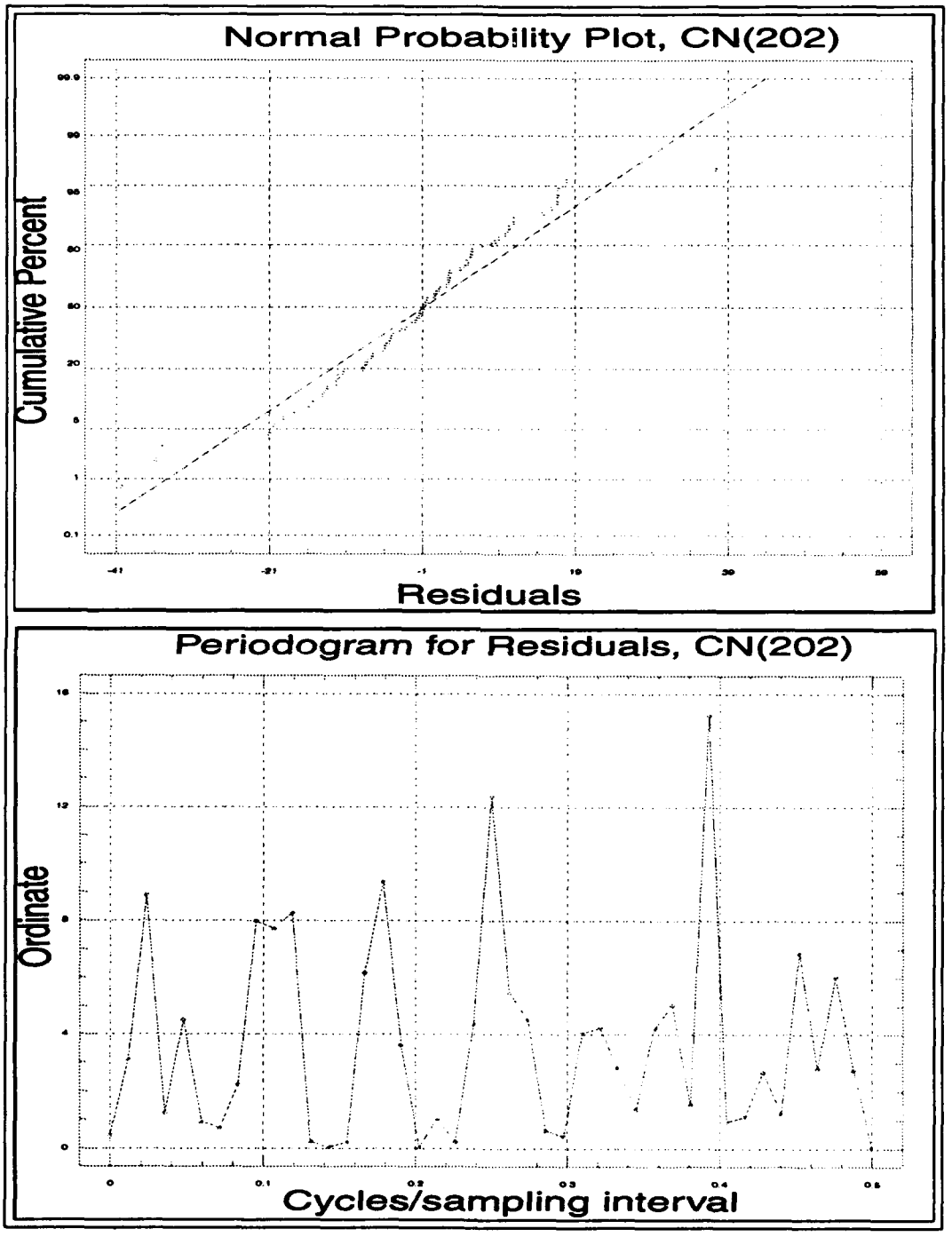

Figure F.88. Cumulative probability plots and periodograms for Model CN(2,0,2).

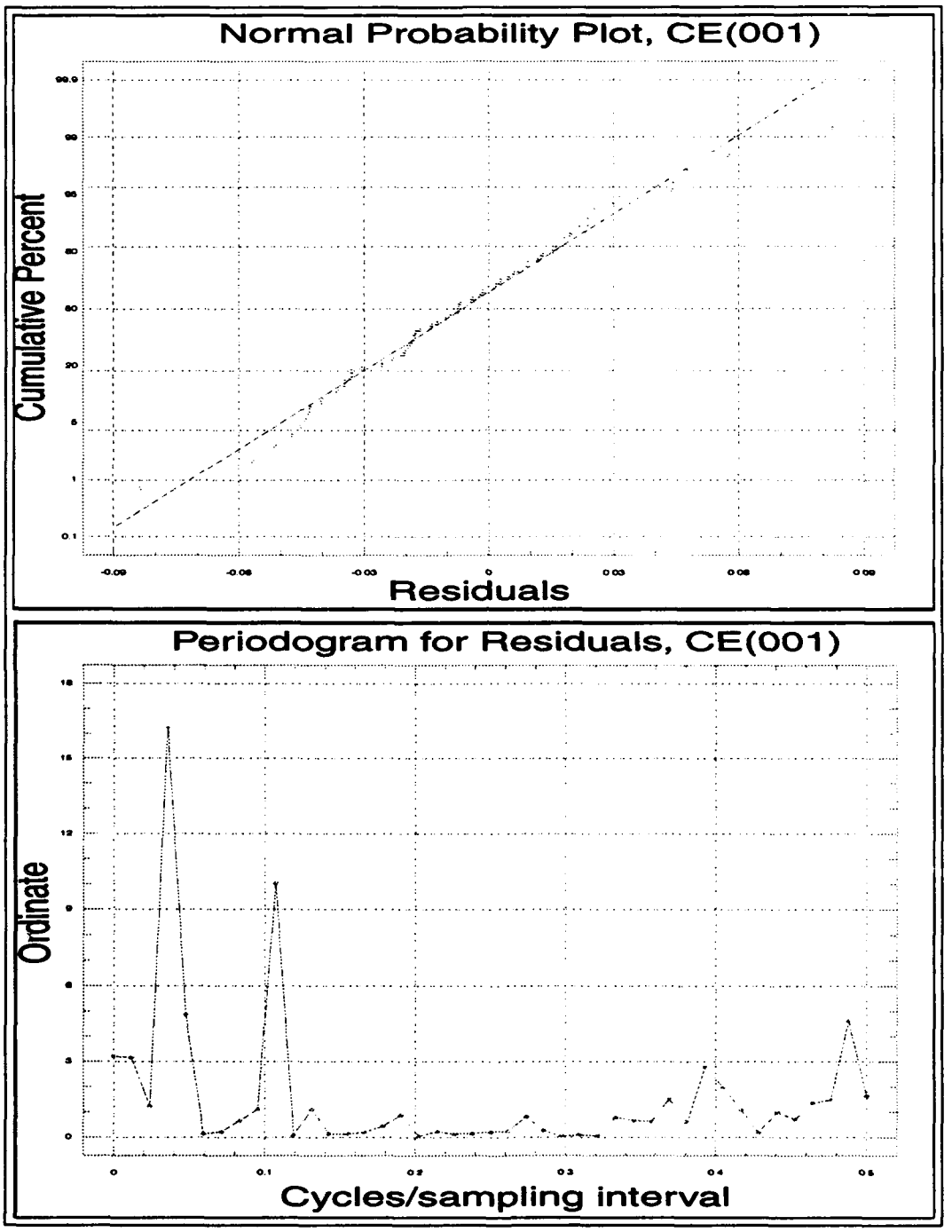

Figure F.89. Cumulative probability plots and periodograms for Model CE(0,0,1).

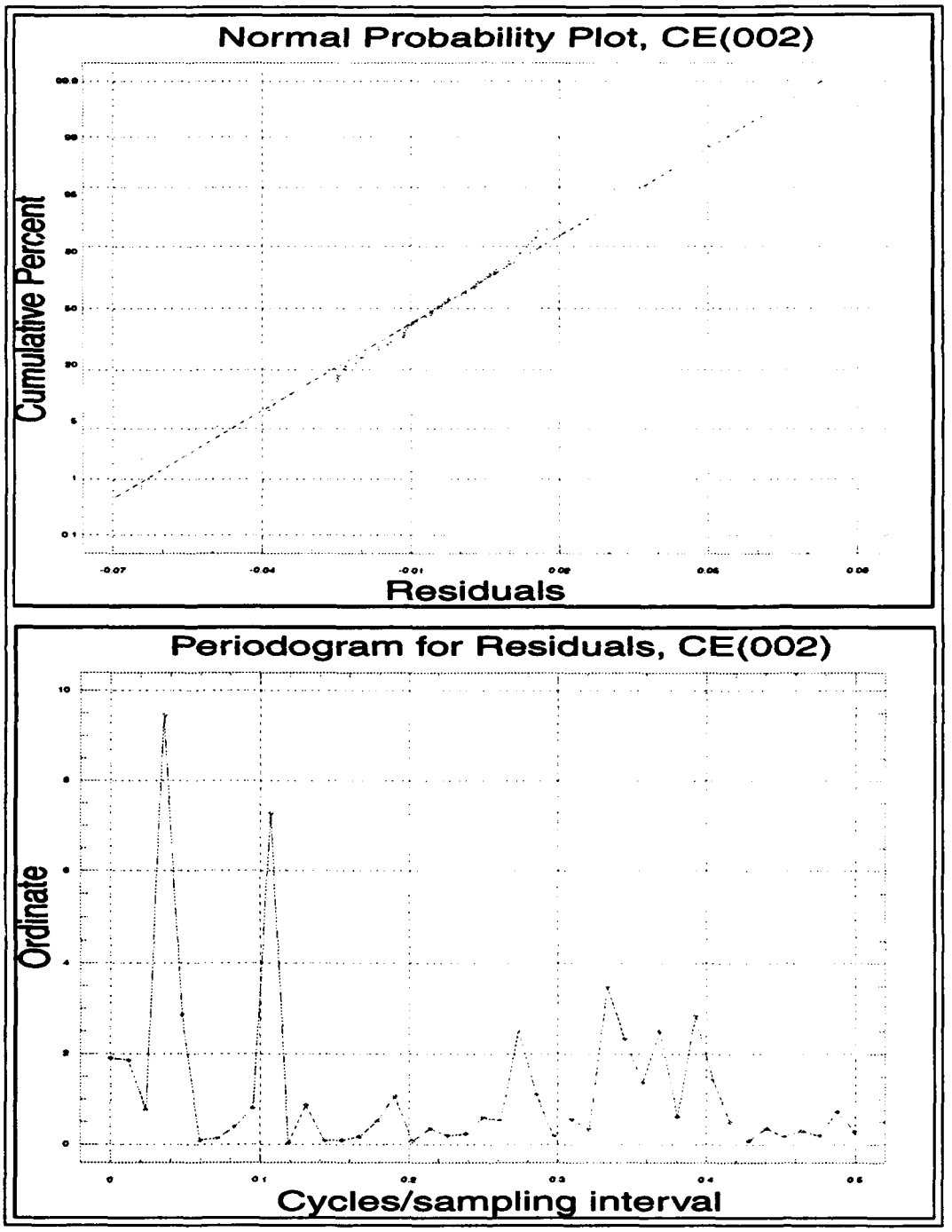

Figure F.90. Cumulative probability plots and periodograms for Model CE(0,0,2).

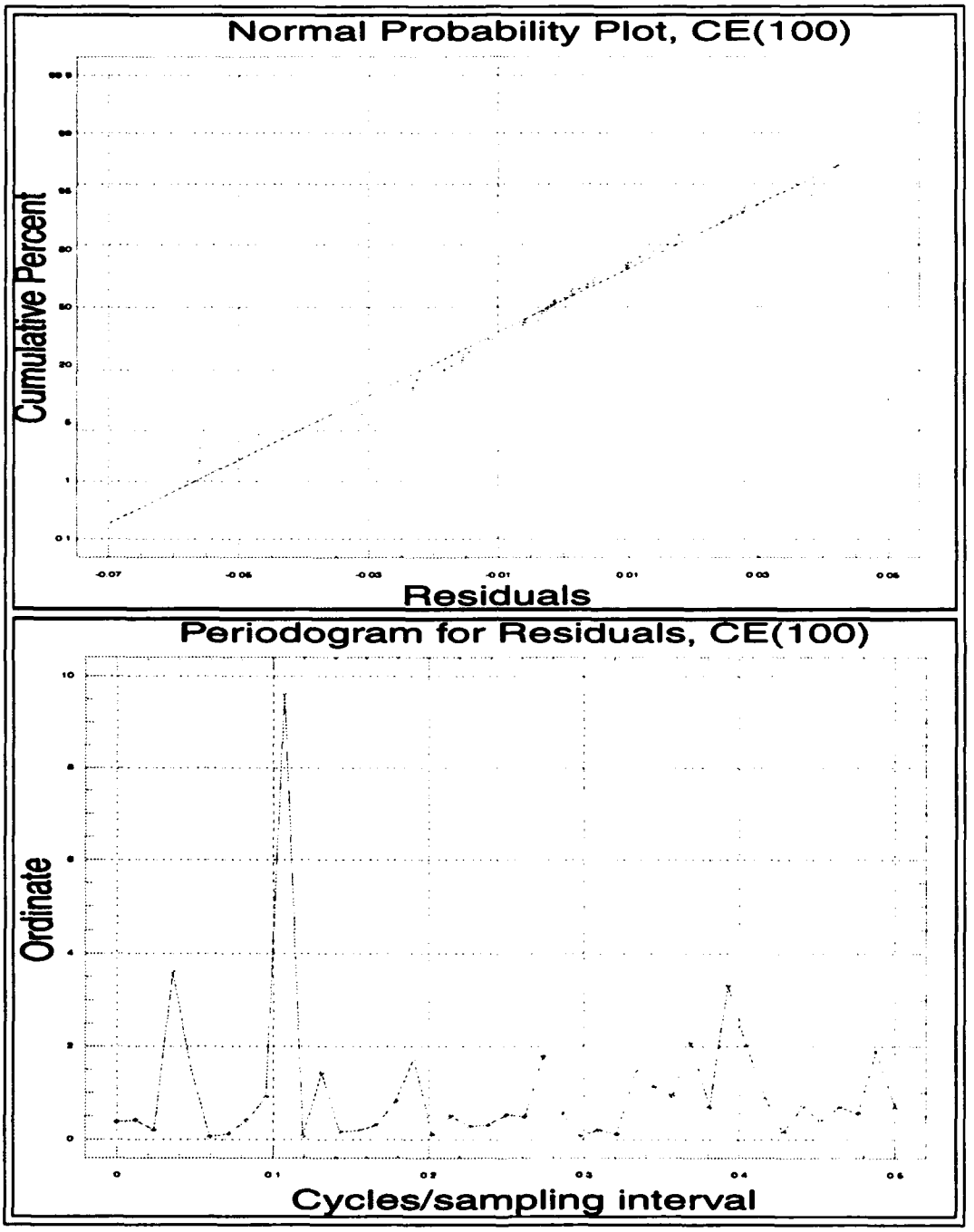

Figure F.91. Cumulative probability plots and periodograms for Model CE(1,0,0).

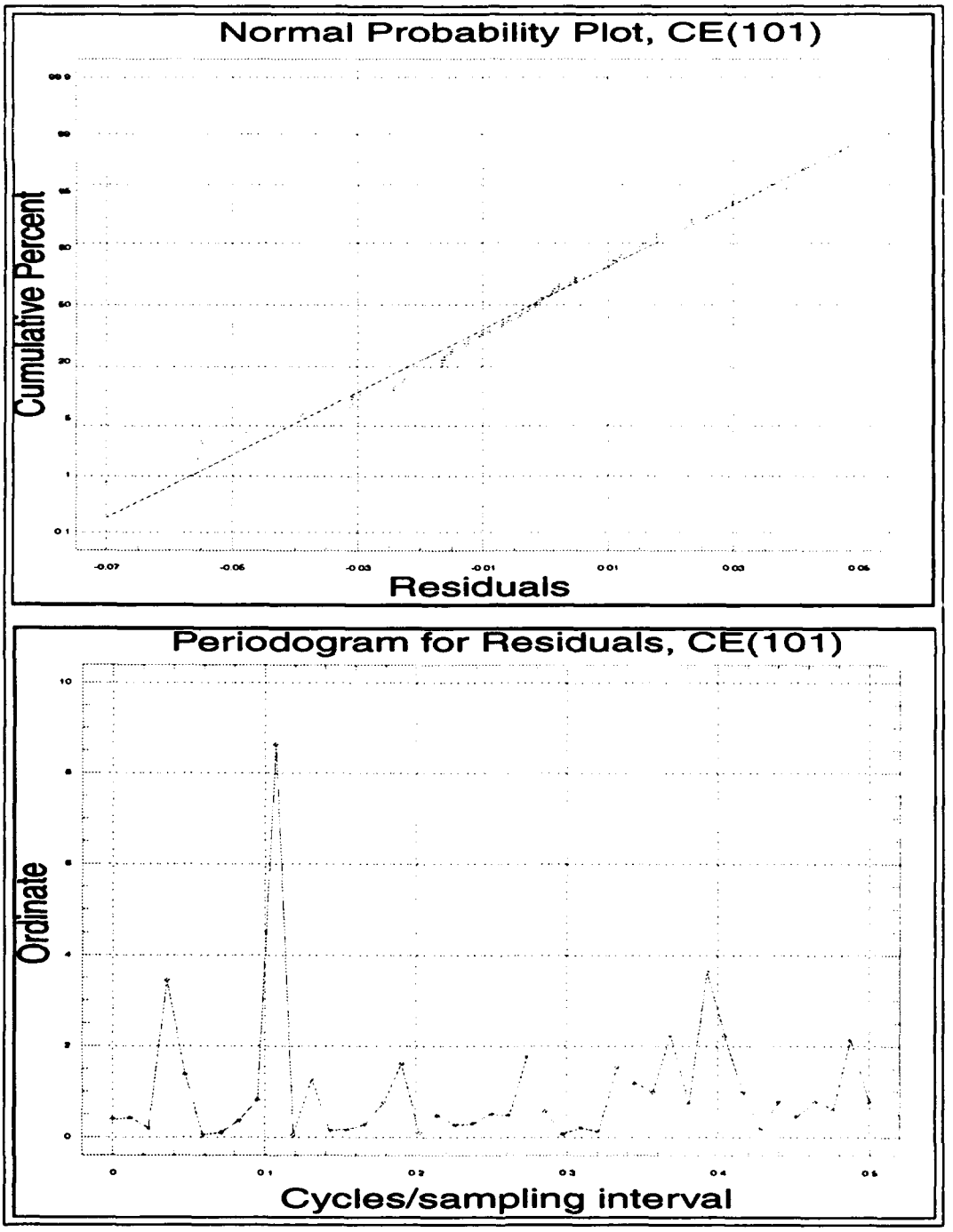

Figure F.92. Cumulative probability plots and periodograms for Model CE(1,0,1).

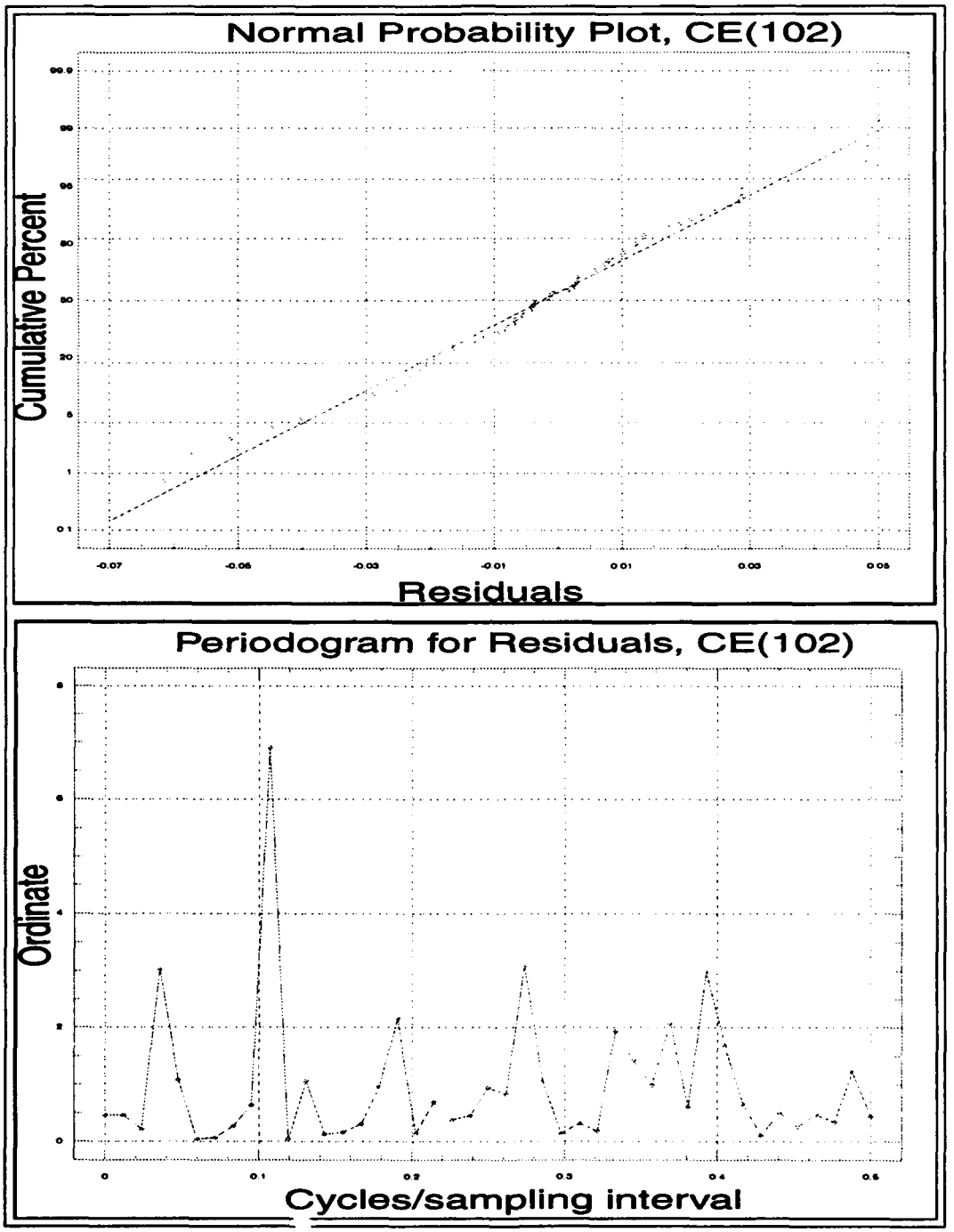

Figure F.93. Cumulative probability plots and periodograms for Model CE(1,0,2).

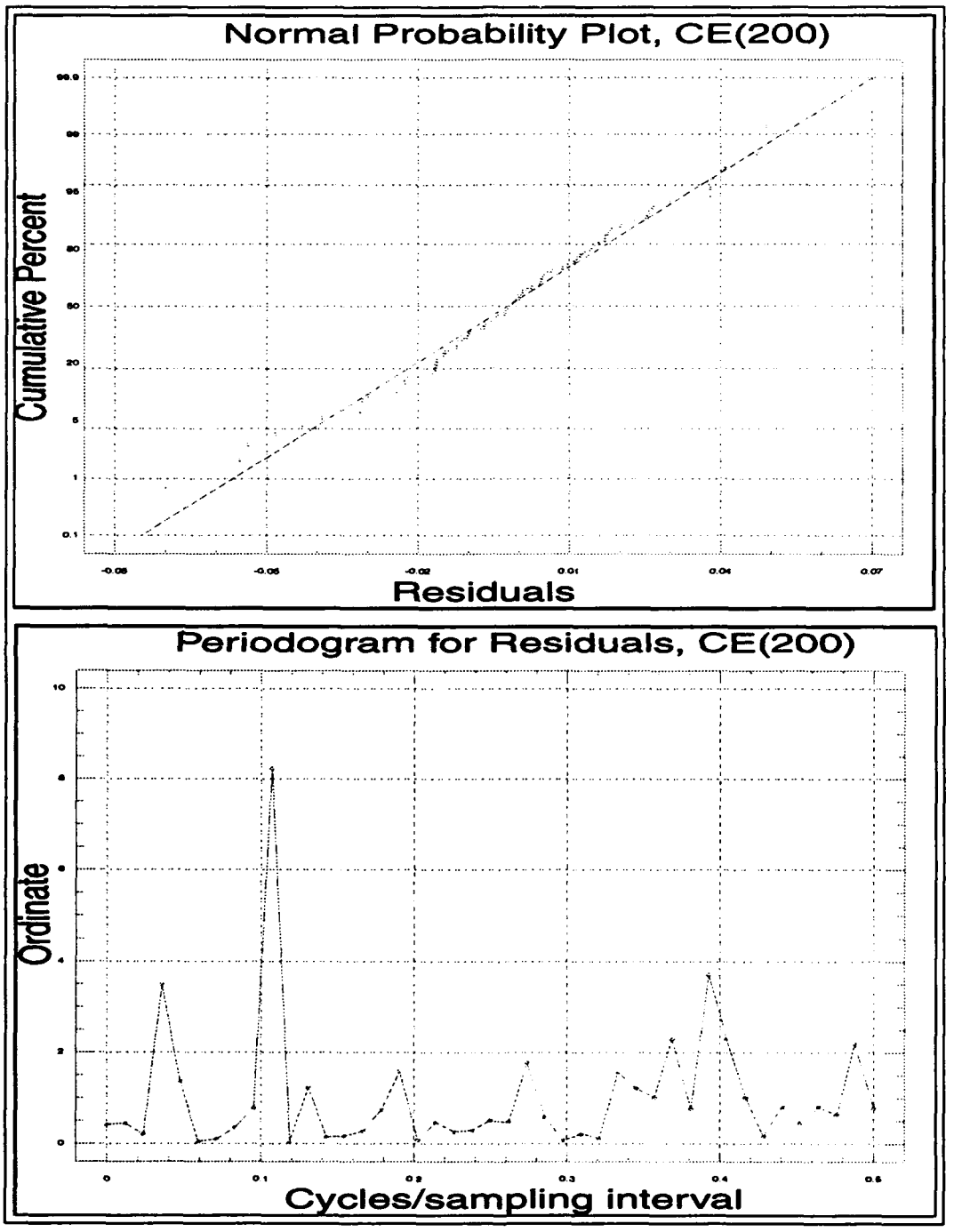

Figure F.94. Cumulative probability plots and periodograms for Model CE(2,0,0).

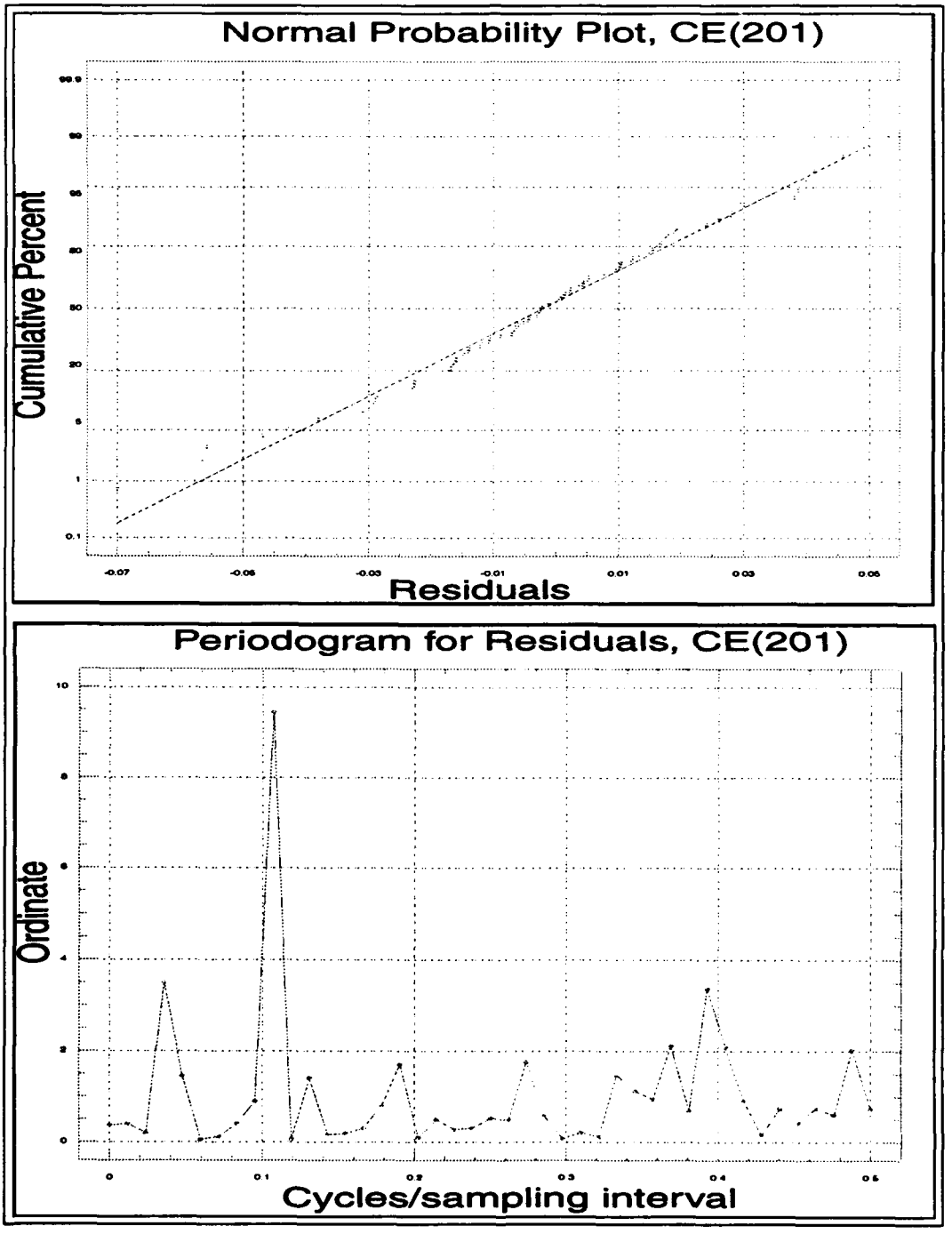

Figure F.95. Cumulative probability plots and periodograms for Model CE(2,0,1).

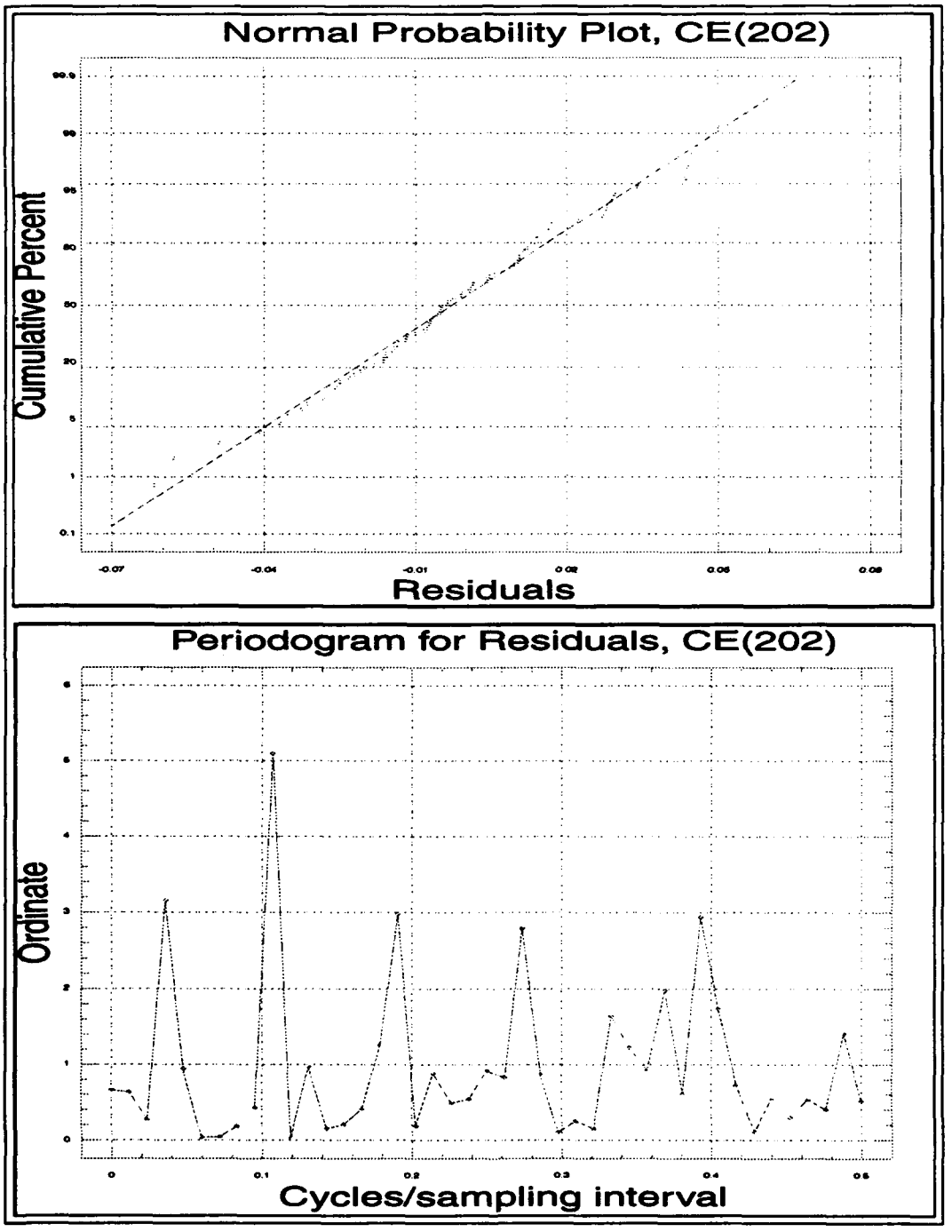

Figure F.96. Cumulative probability plots and periodograms for Model CE(2,0,2).

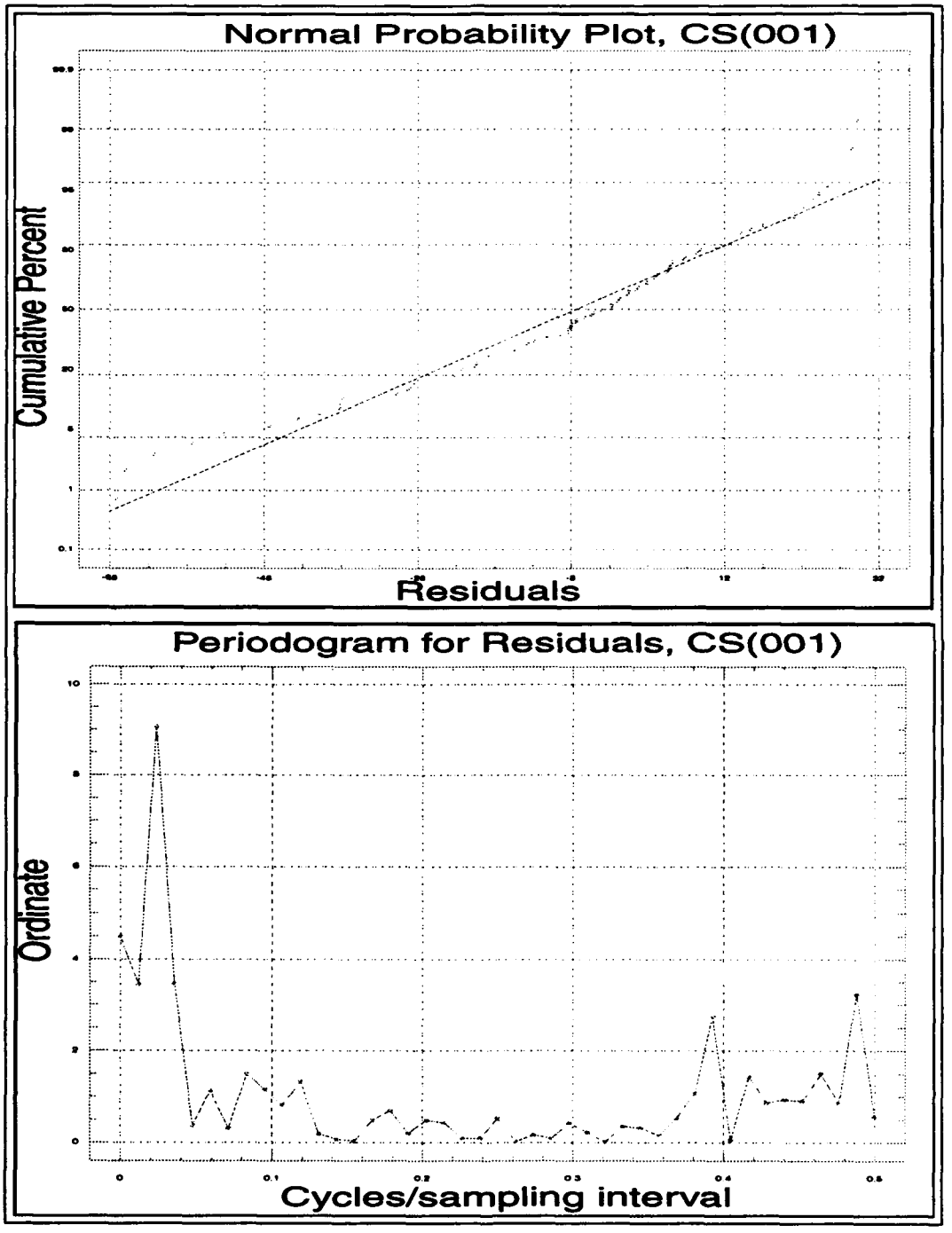

Figure F.97. Cumulative probability plots and periodograms for Model CS(0,0,1).

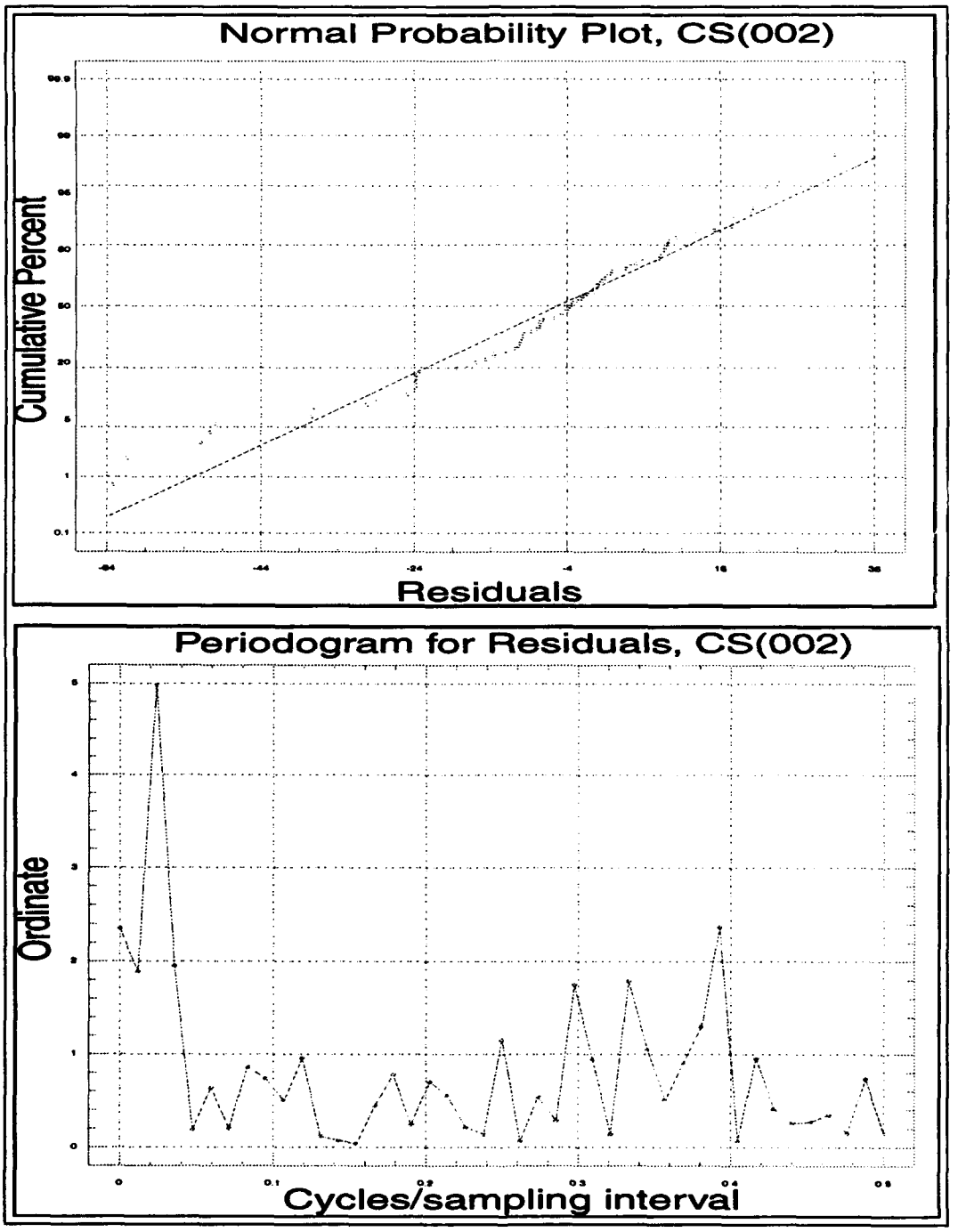

Figure F.98. Cumulative probability plots and periodograms for Model CS(0,0,2).

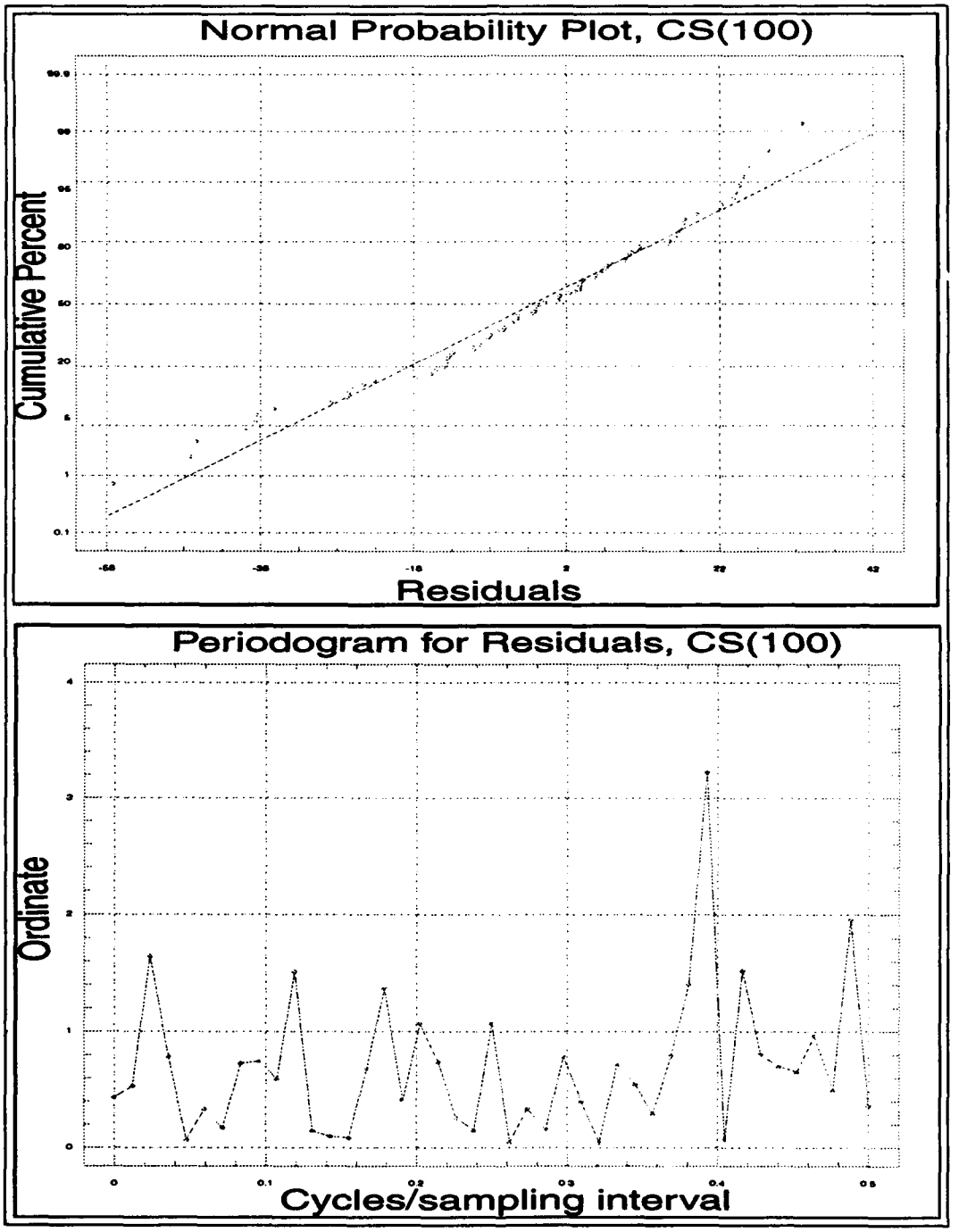

Figure F.99. Cumulative probability plots and periodograms for Model CS(1,0,0).

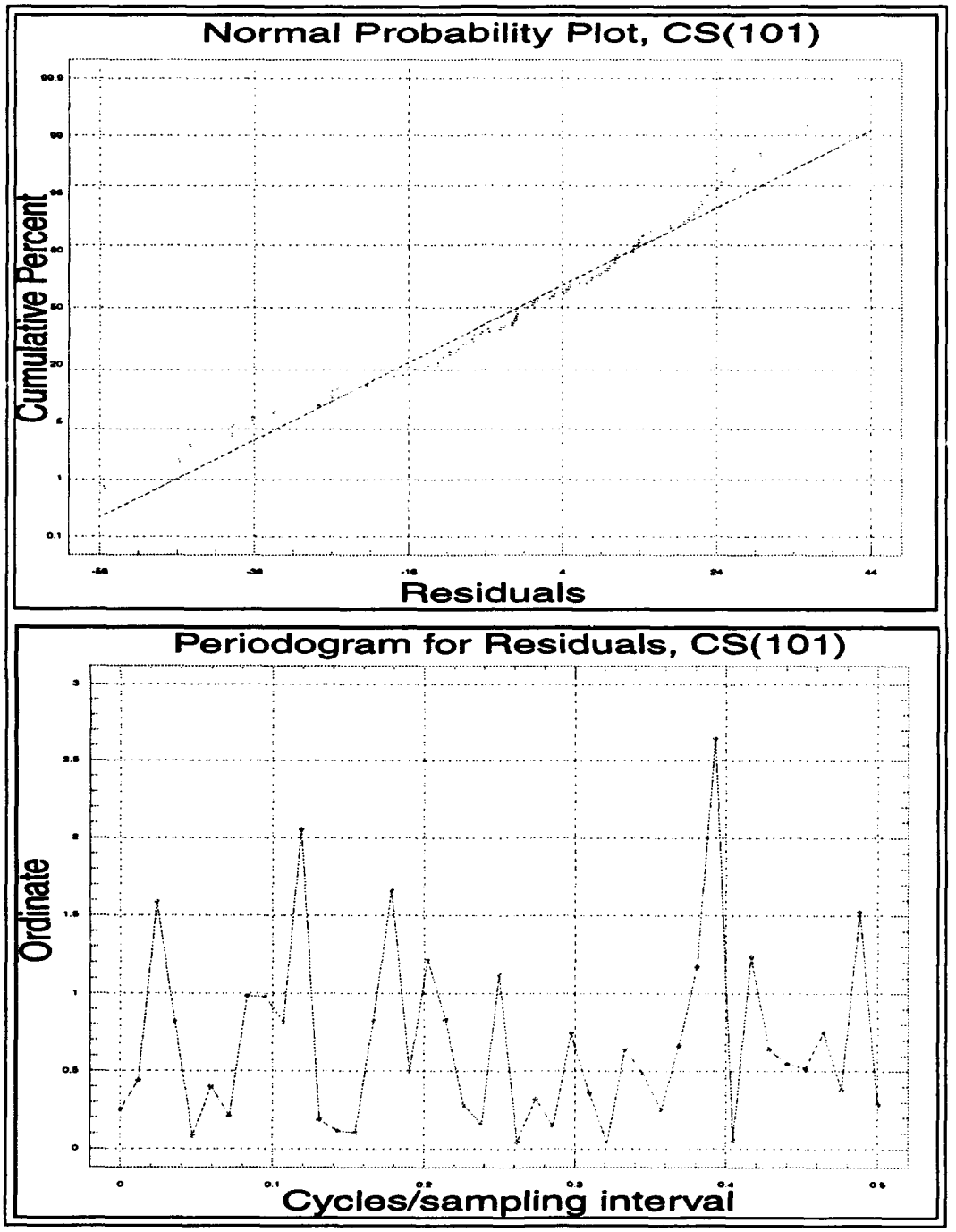

Figure F.100. Cumulative probability plots and periodograms for Model CS(1,0,1).

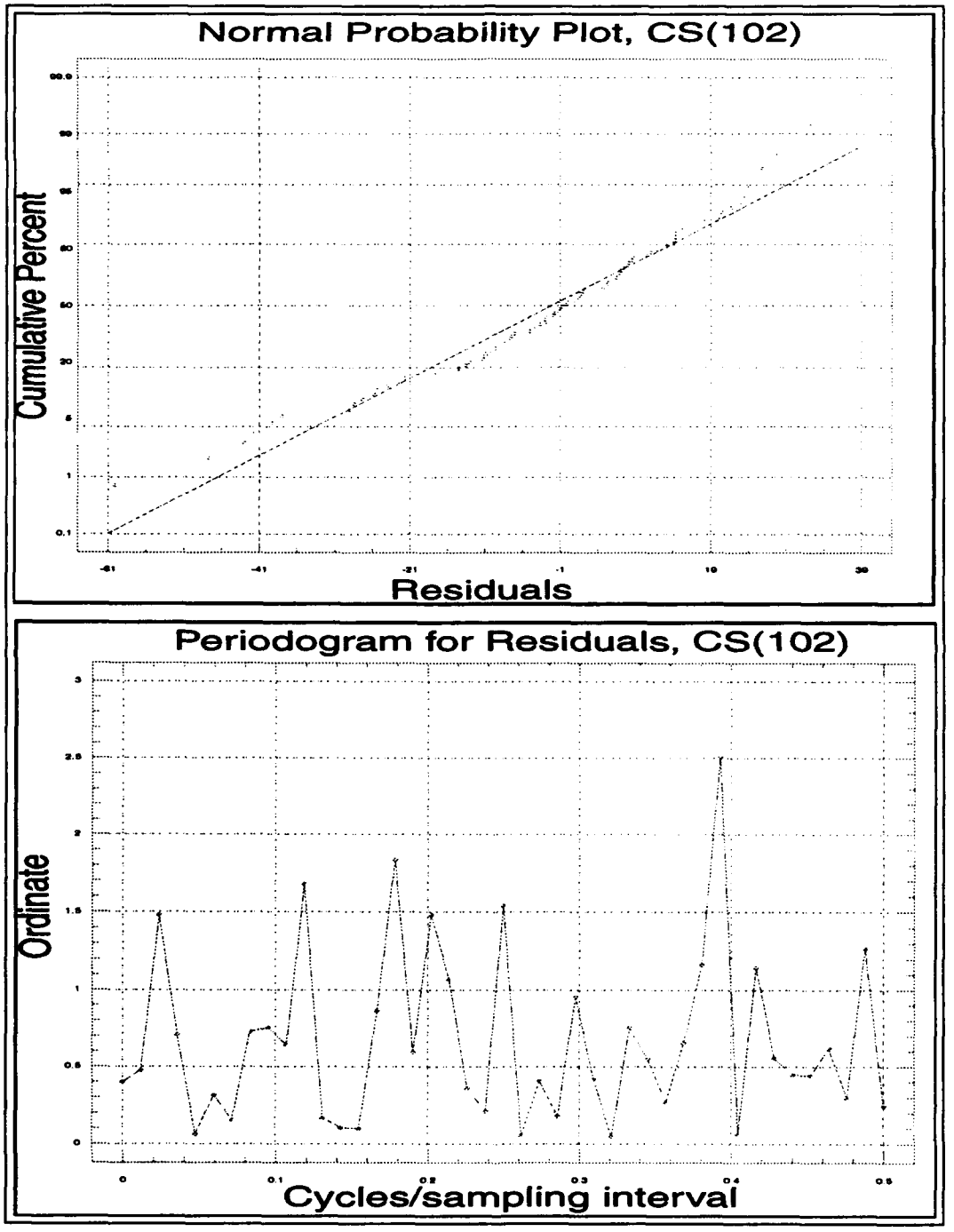

Figure F.101. Cumulative probability plots and periodograms for Model CS(1,0,2).

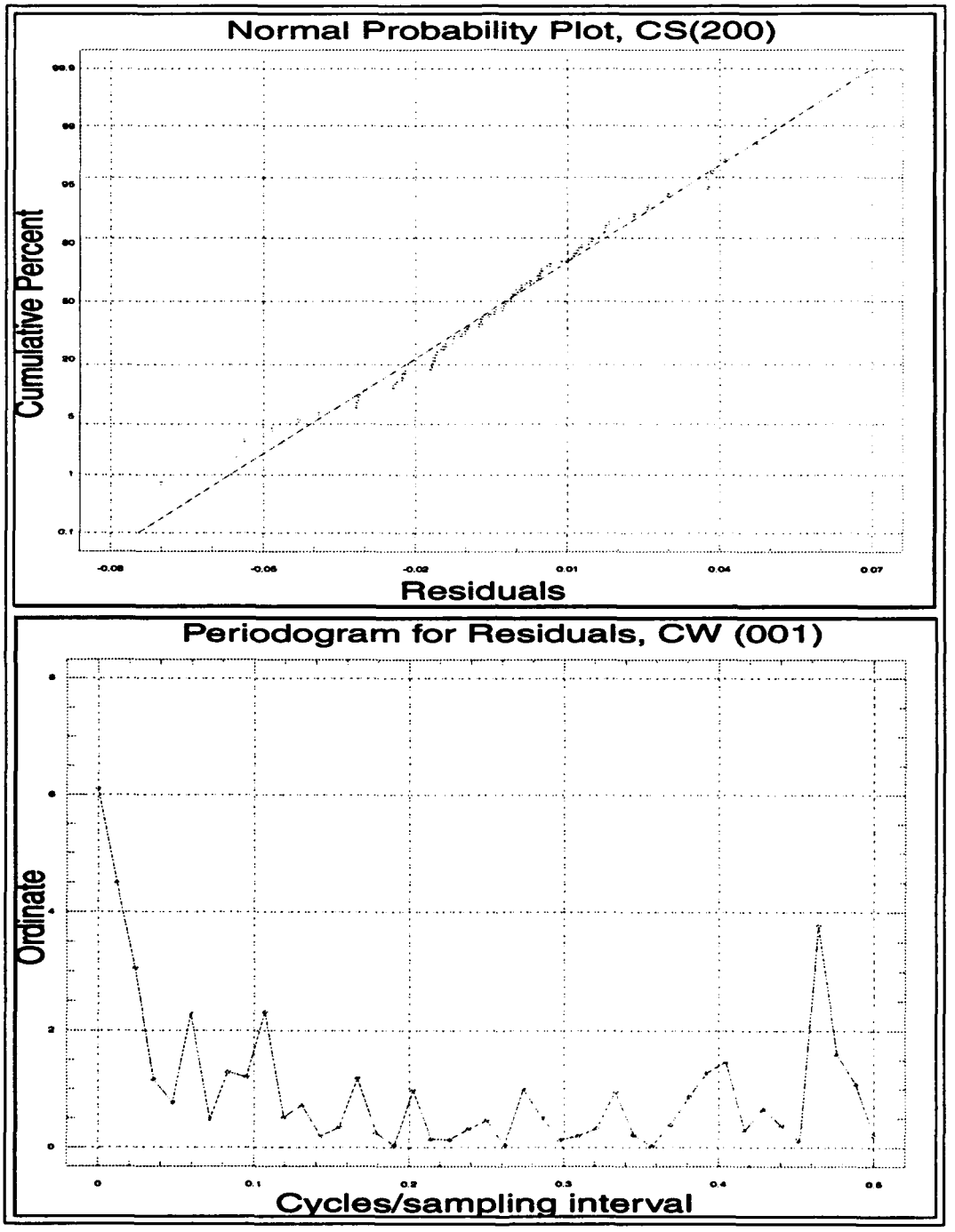

Figure F.102. Cumulative probability plots and periodograms for Model CS(2,0,0).

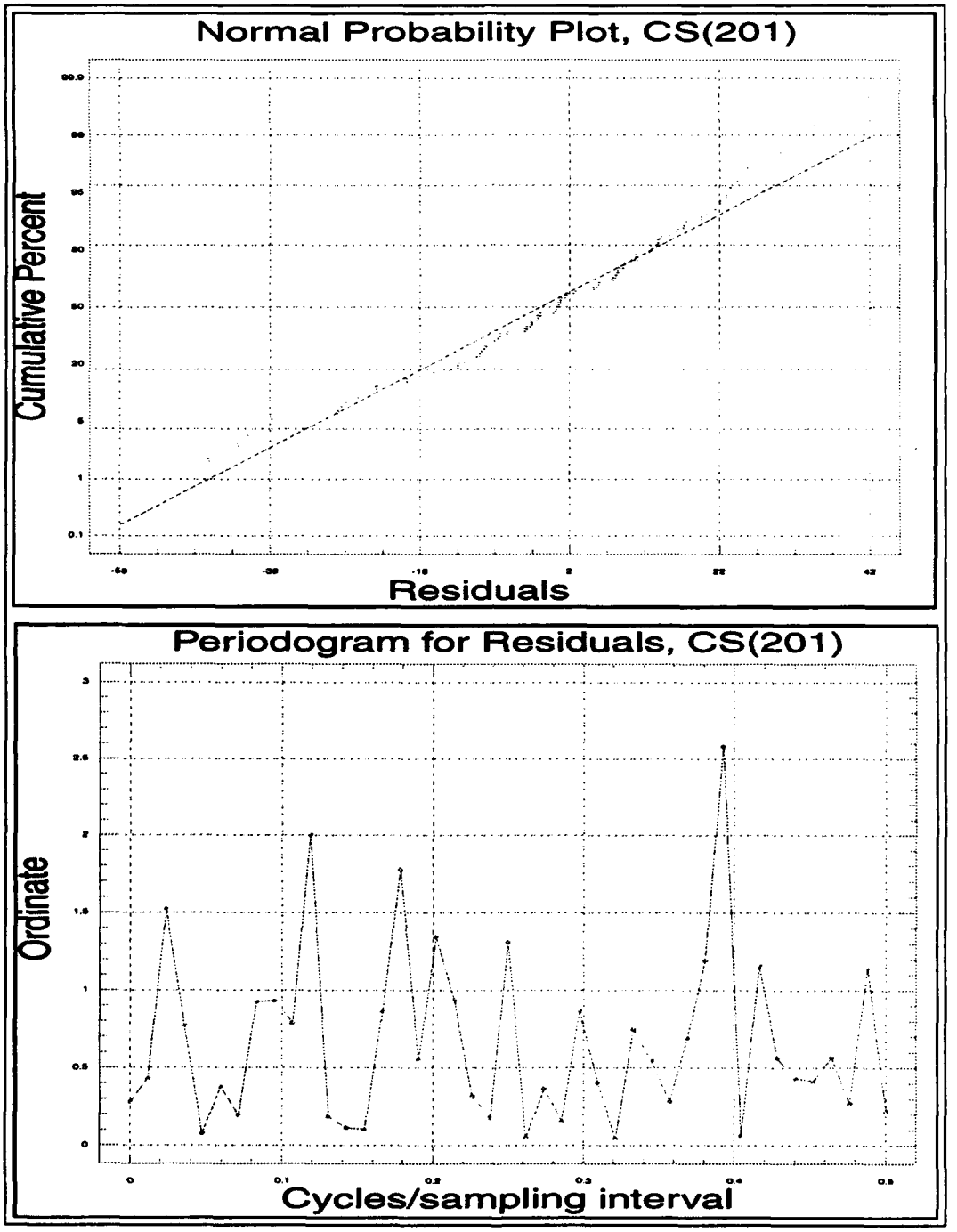

Figure F.103. Cumulative probability plots and periodograms for Model CS(2,0,1).

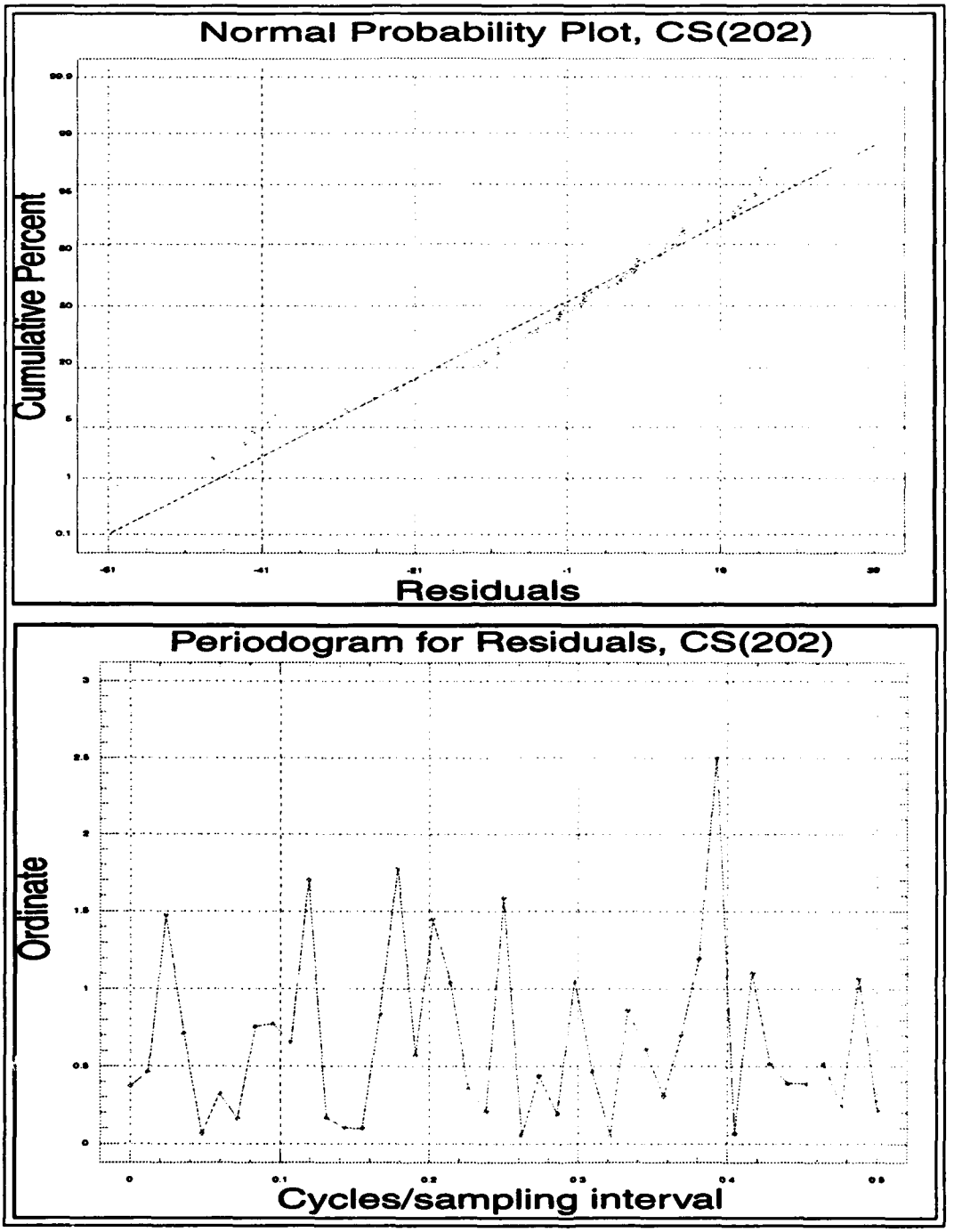

Figure F.104. Cumulative probability plots and periodograms for Model CS(2,0,2).

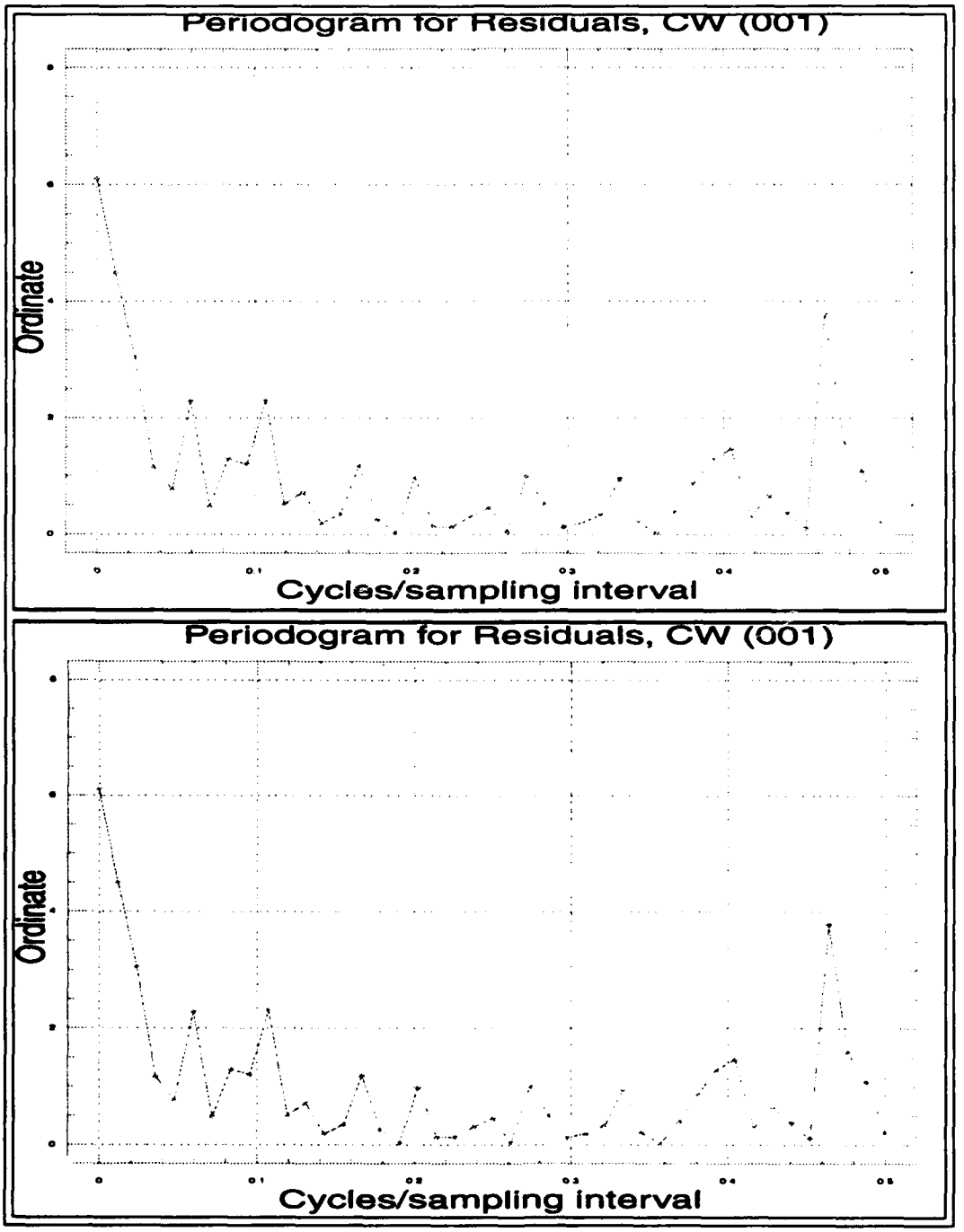

Figure F.105. Cumulative probability plots and periodograms for Model CW(0,0,1).

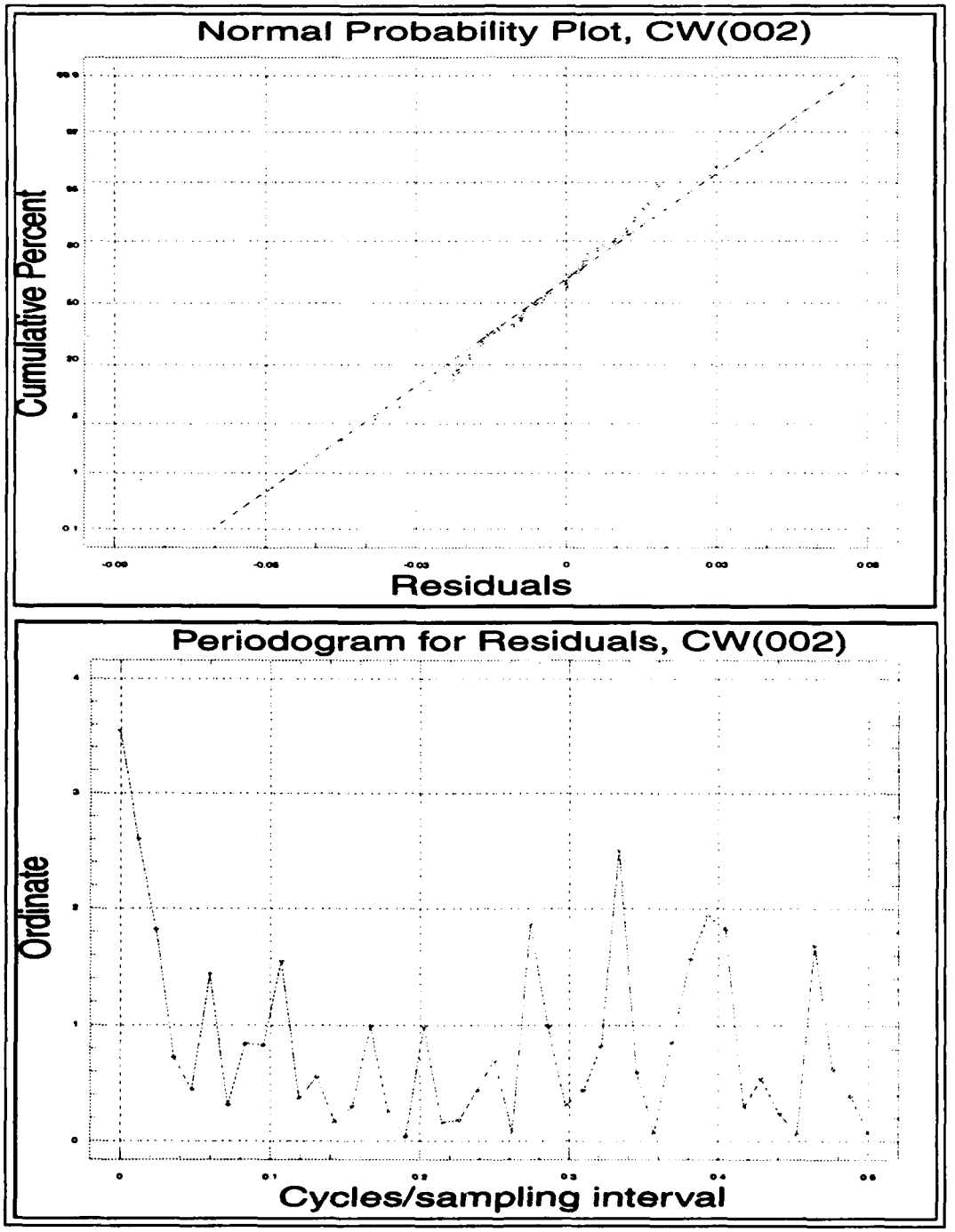

Figure F.106. Cumulative probability plots and periodograms for Model CW(0,0,2).

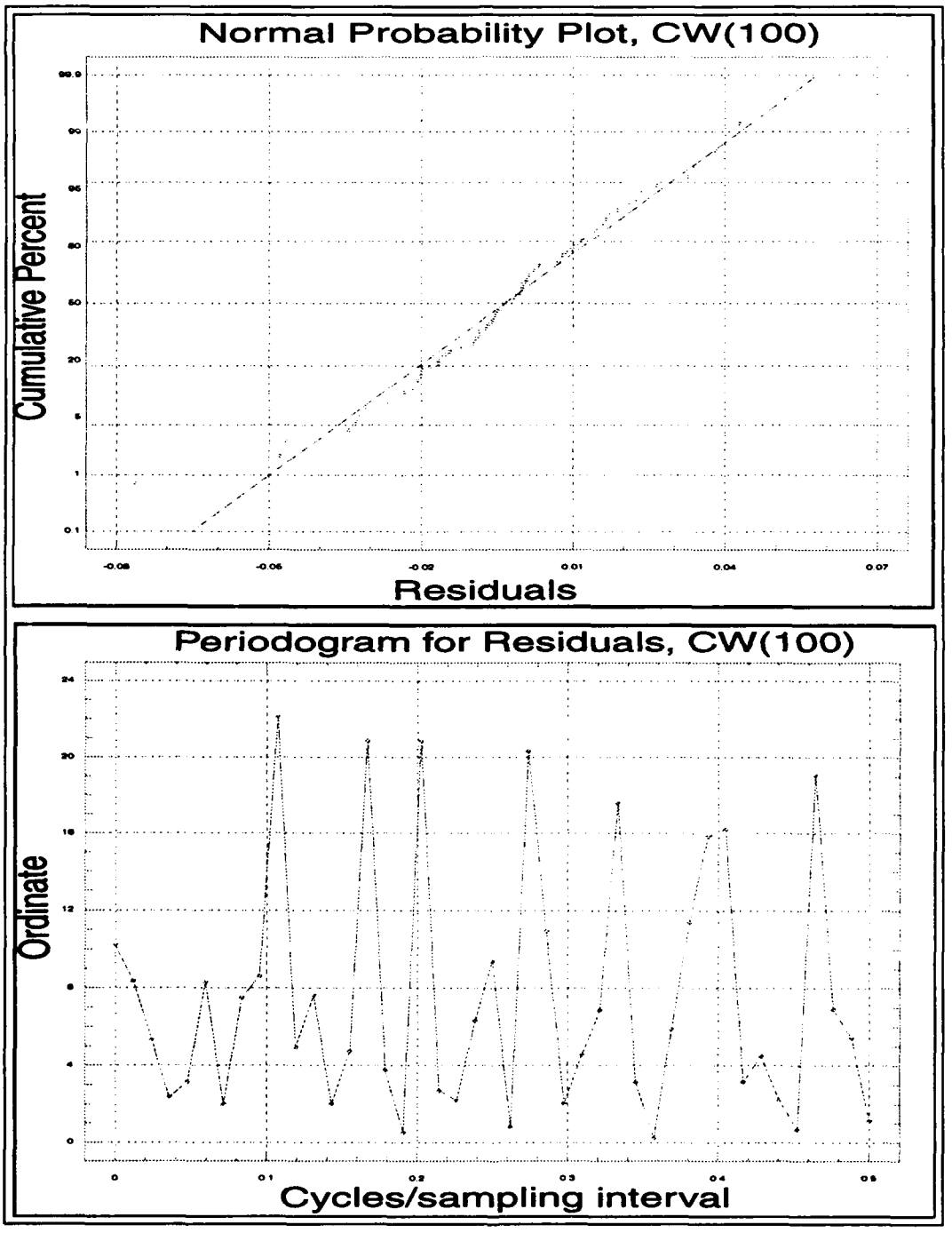

Figure F.107. Cumulative probability plots and periodograms for Model CW(1,0,0).

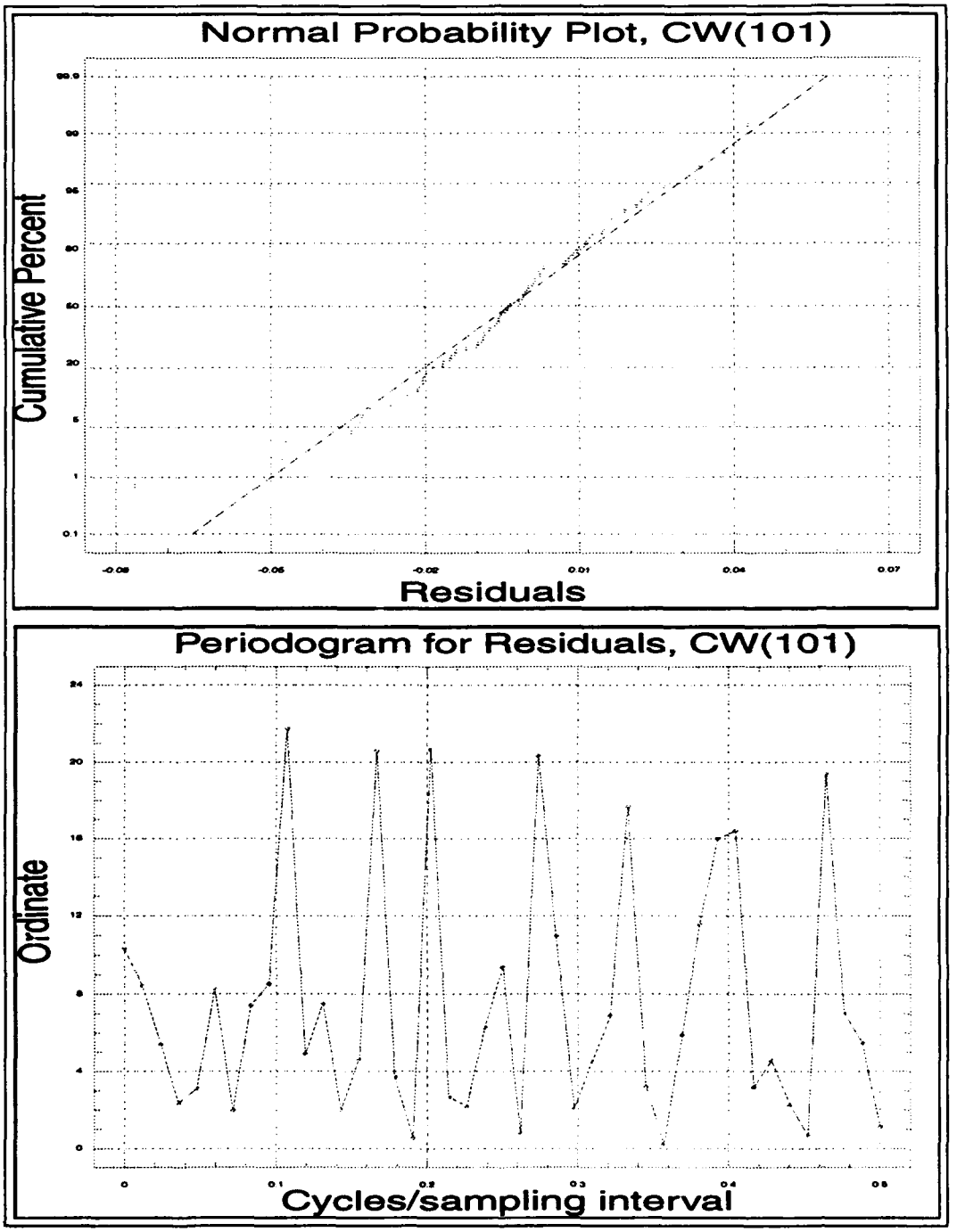

Figure F.108. Cumulative probability plots and periodograms for Model CW(1,0,1).

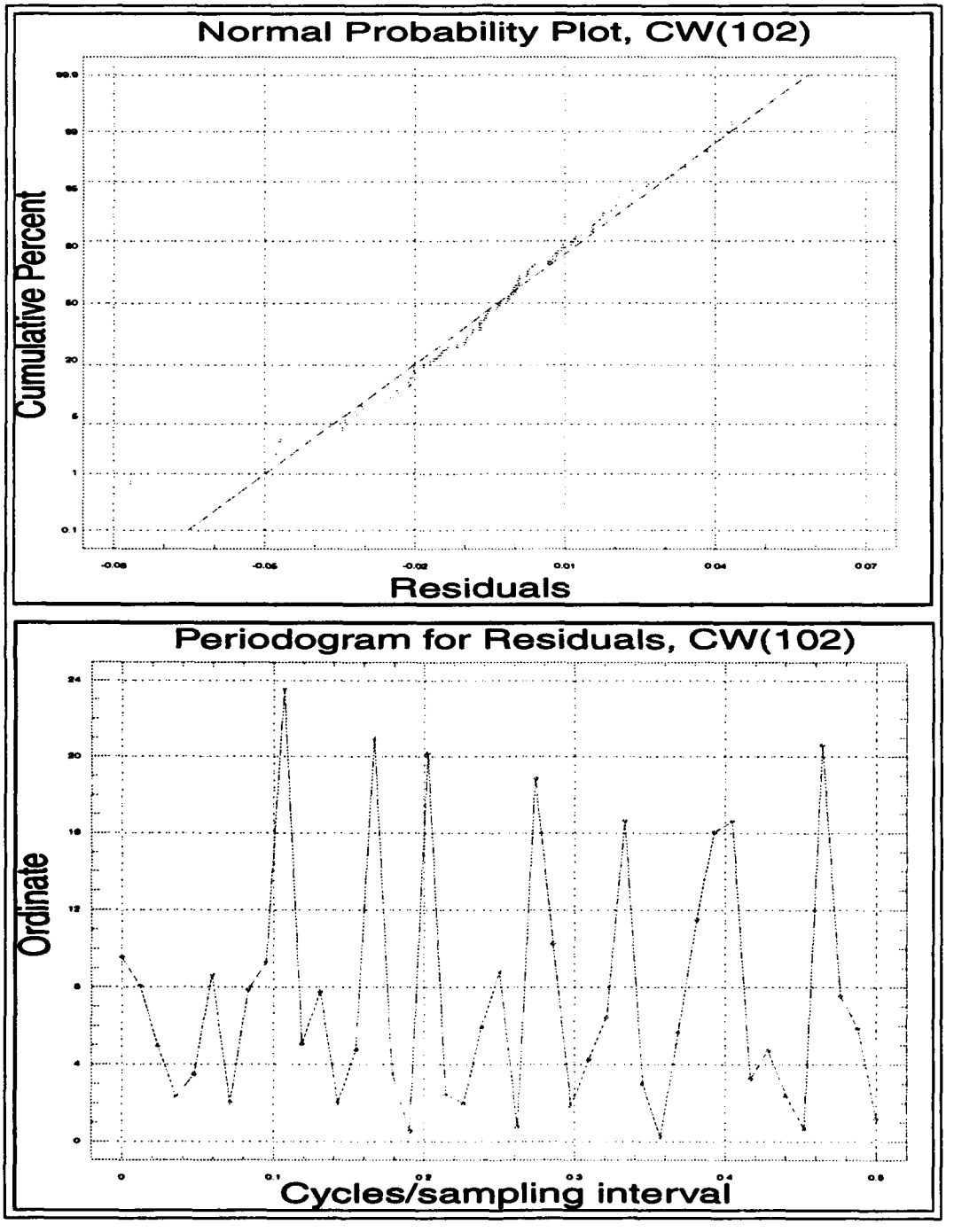

Figure F.109. Cumulative probability plots and periodograms for Model CW(1,0,2).

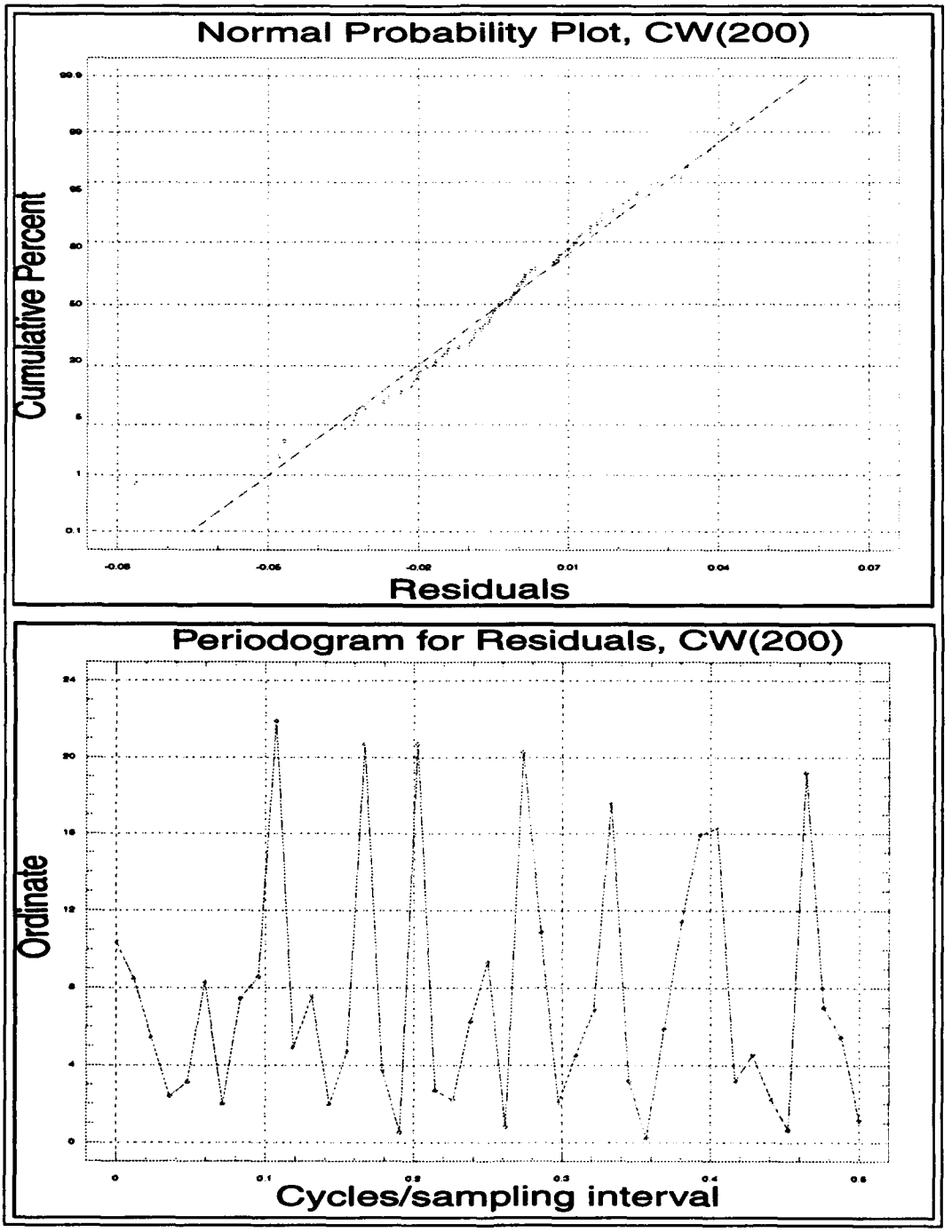

Figure F.110. Cumulative probability plots and periodograms for Model CW(2,0,0).

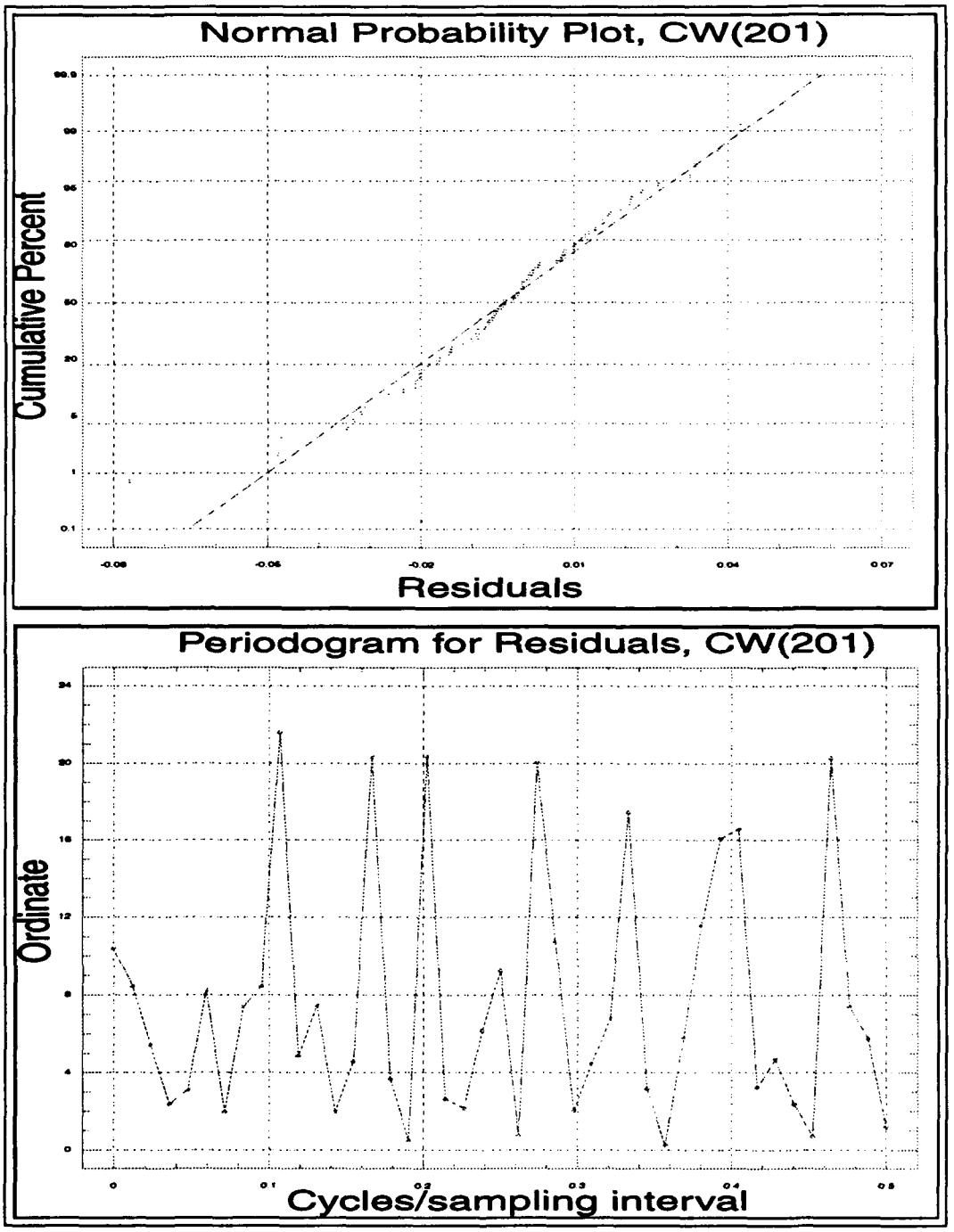

Figure F.111. Cumulative probability plots and periodograms for Model CW(2,0,1).

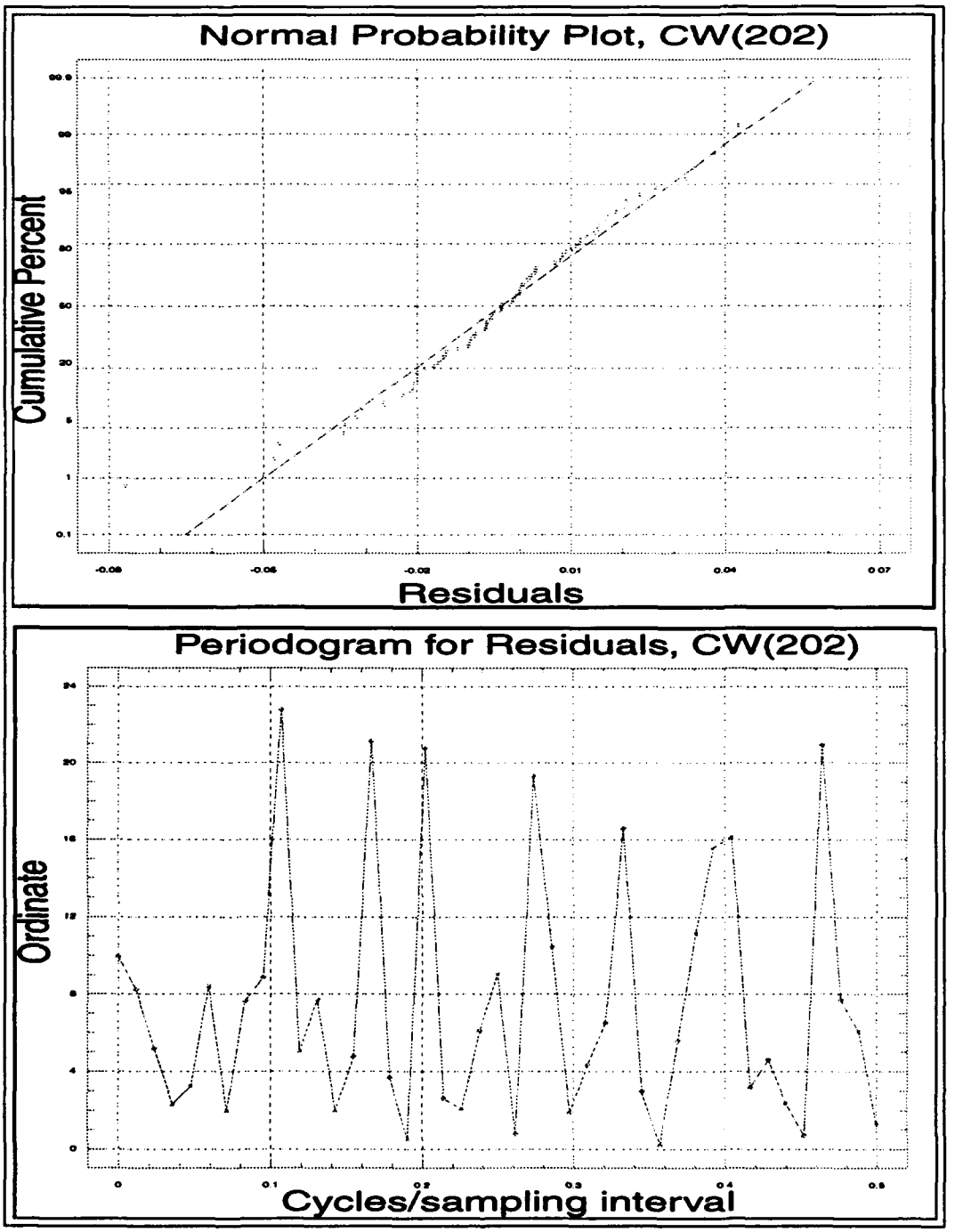

Figure F.112. Cumulative probability plots and periodograms for Model CW(2,0,2).

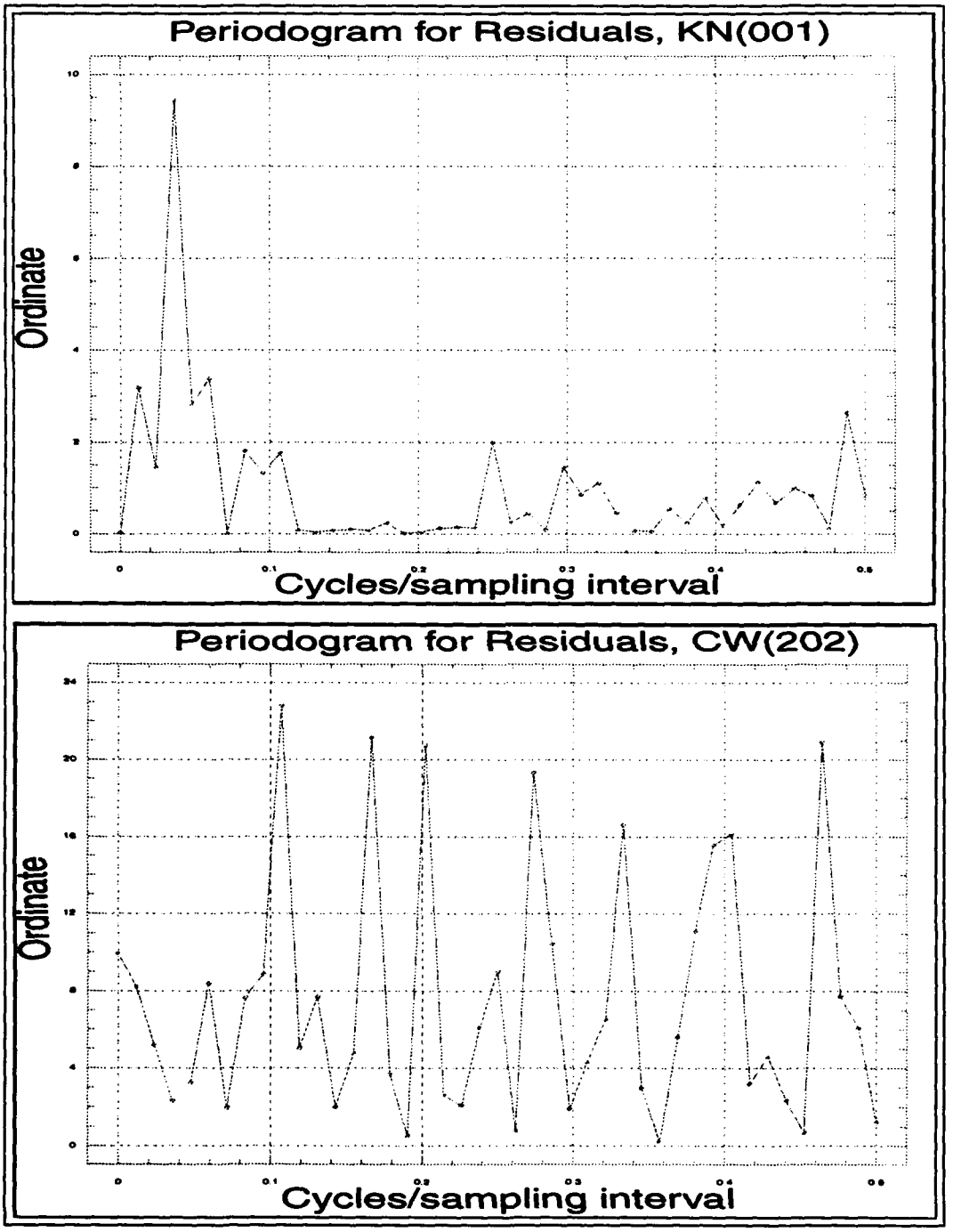

Figure F.113. Cumulative probability plots and periodograms for Model KN(0,0,1).

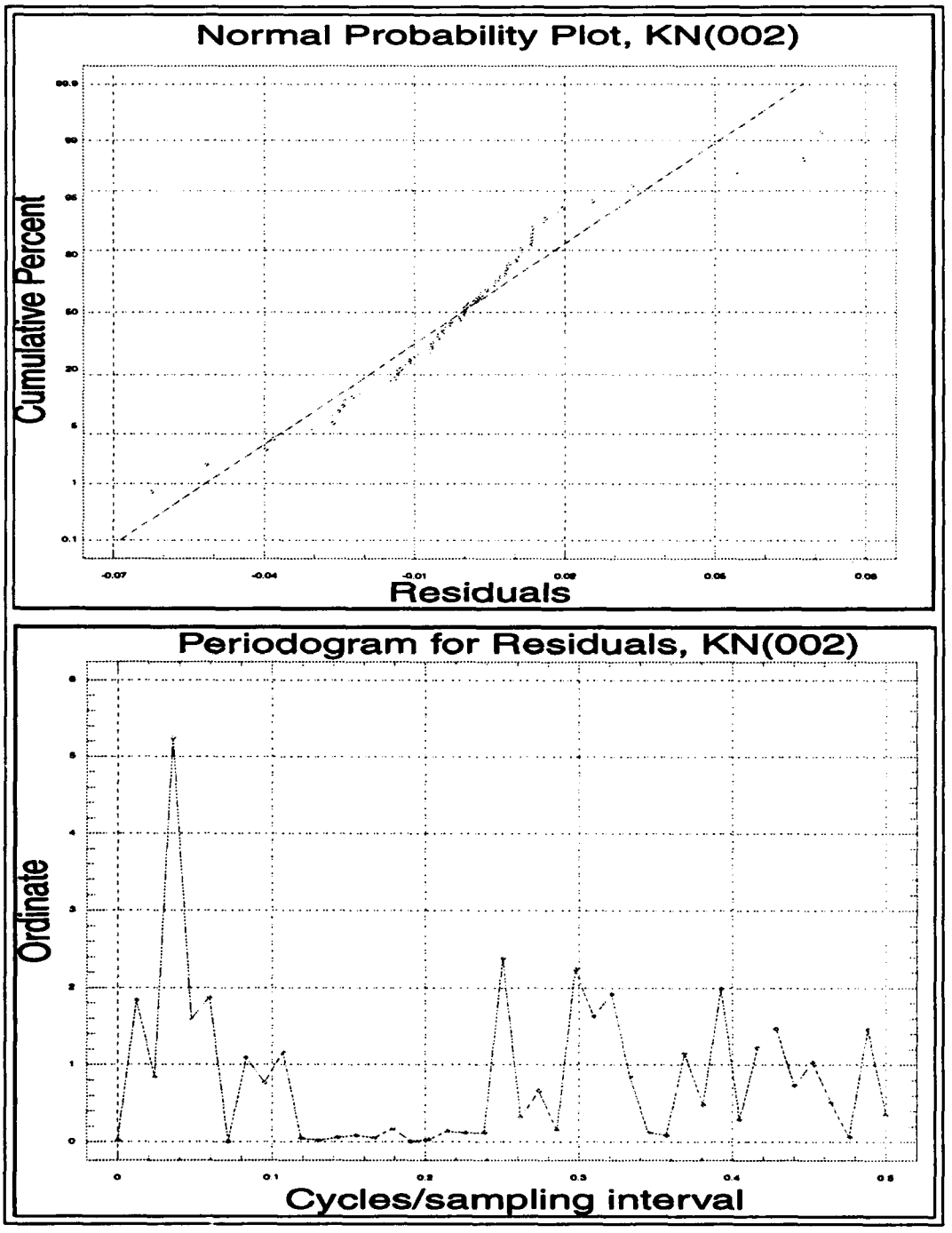

Figure F.114. Cumulative probability plots and periodograms for Model KN(0,0,2).

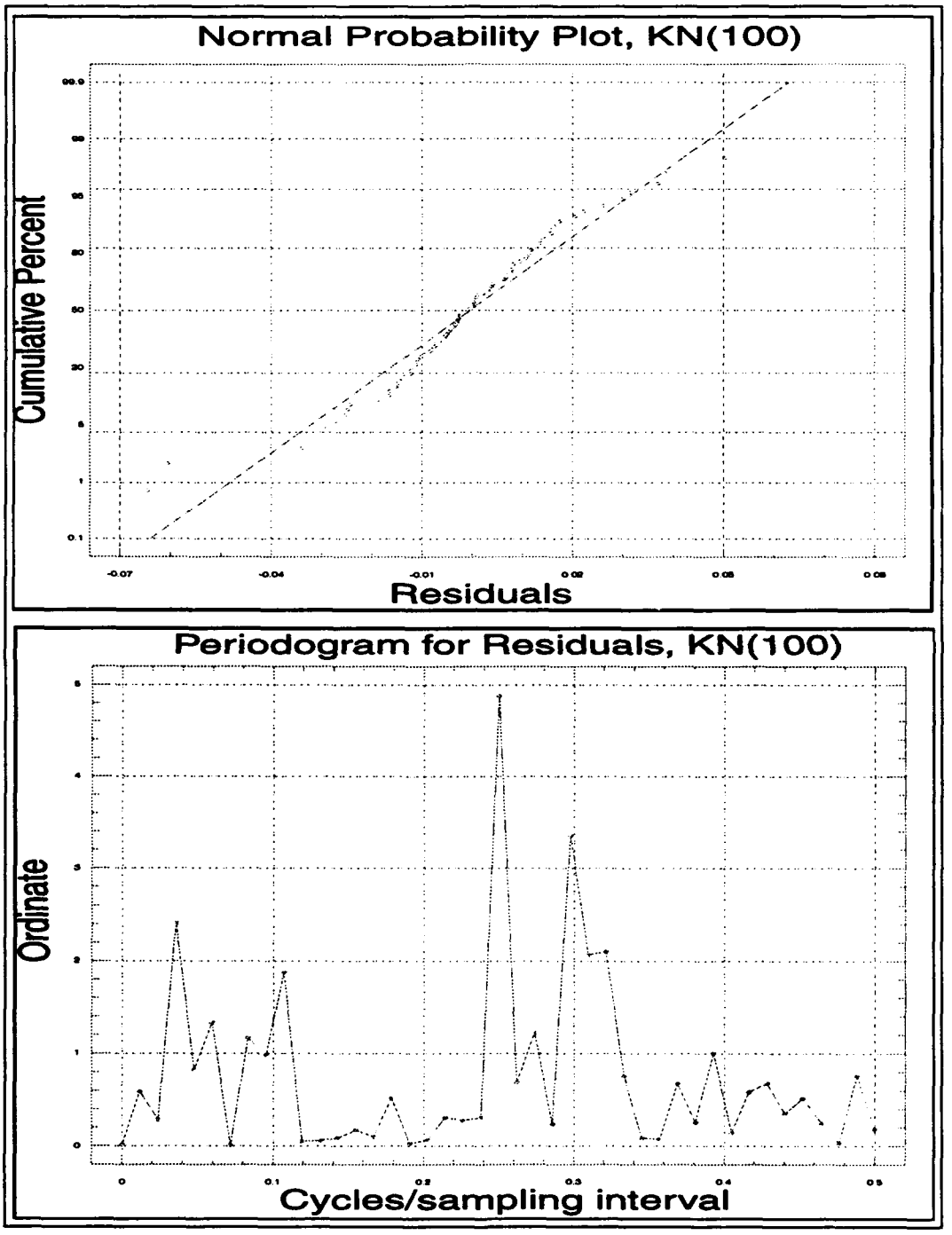

Figure F.115. Cumulative probability plots and periodograms for Model KN(1,0,0).

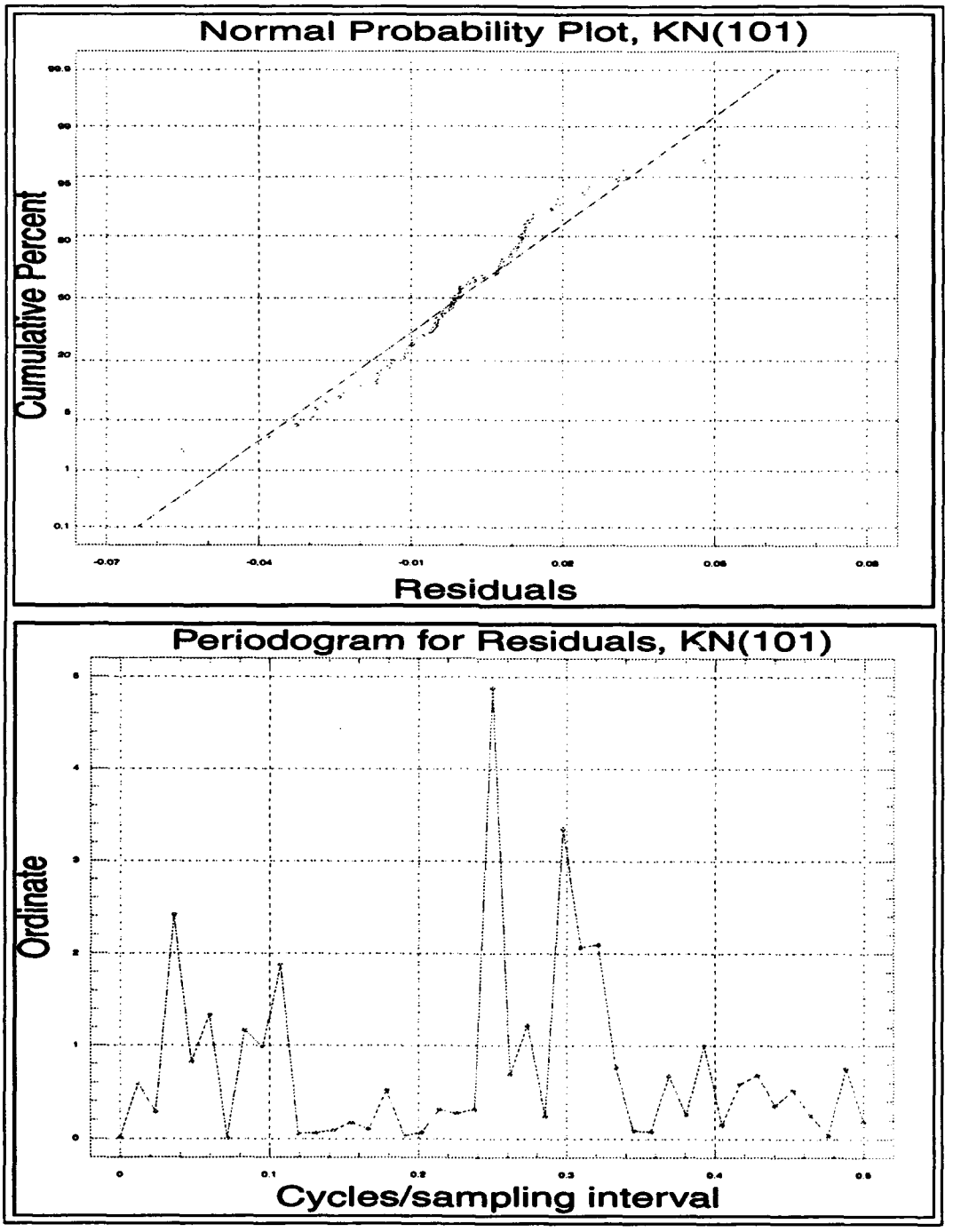

Figure F.116. Cumulative probability plots and periodograms for Model KN(1,0,1).

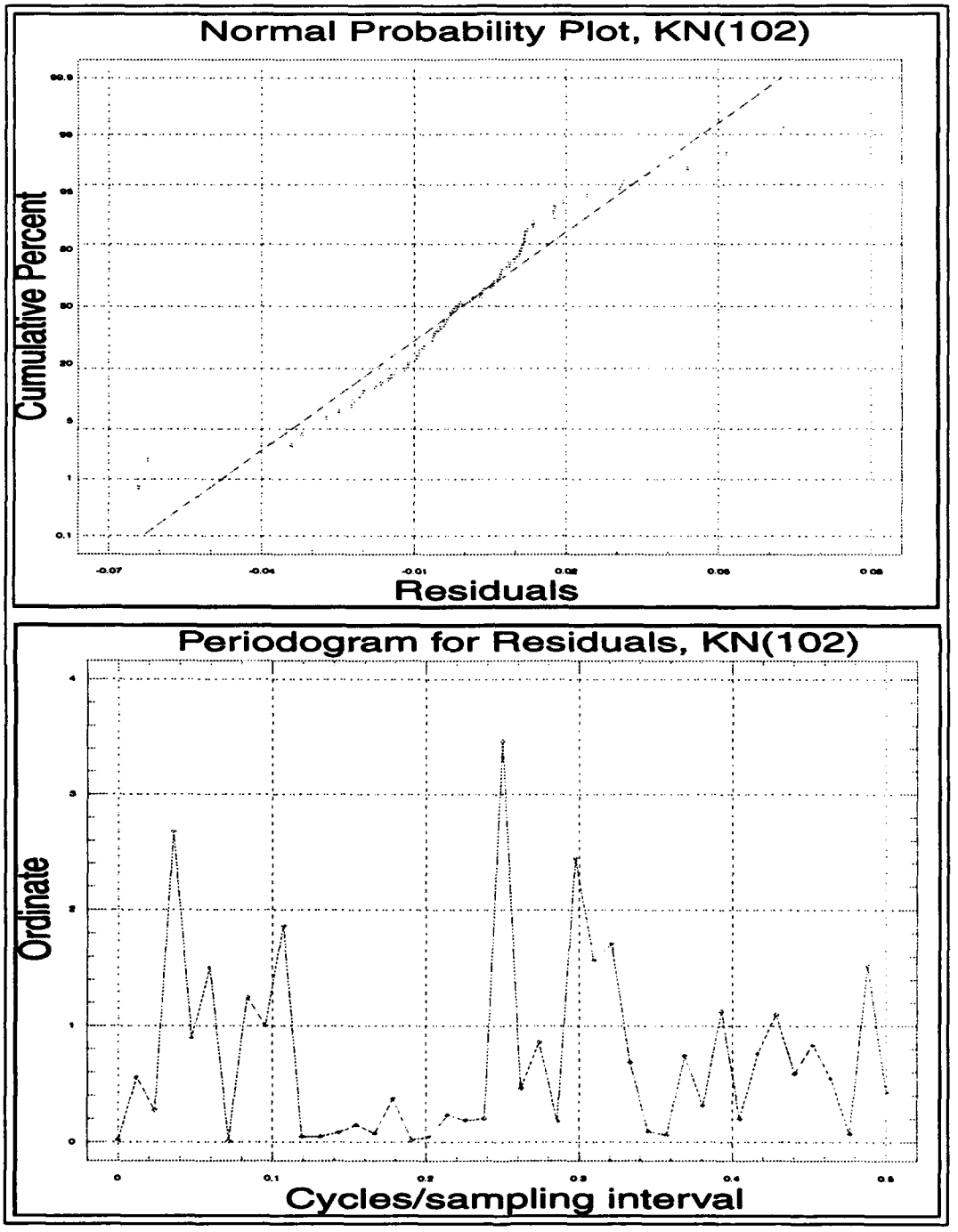

Figure F.117. Cumulative probability plots and periodograms for Model KN(1,0,2).

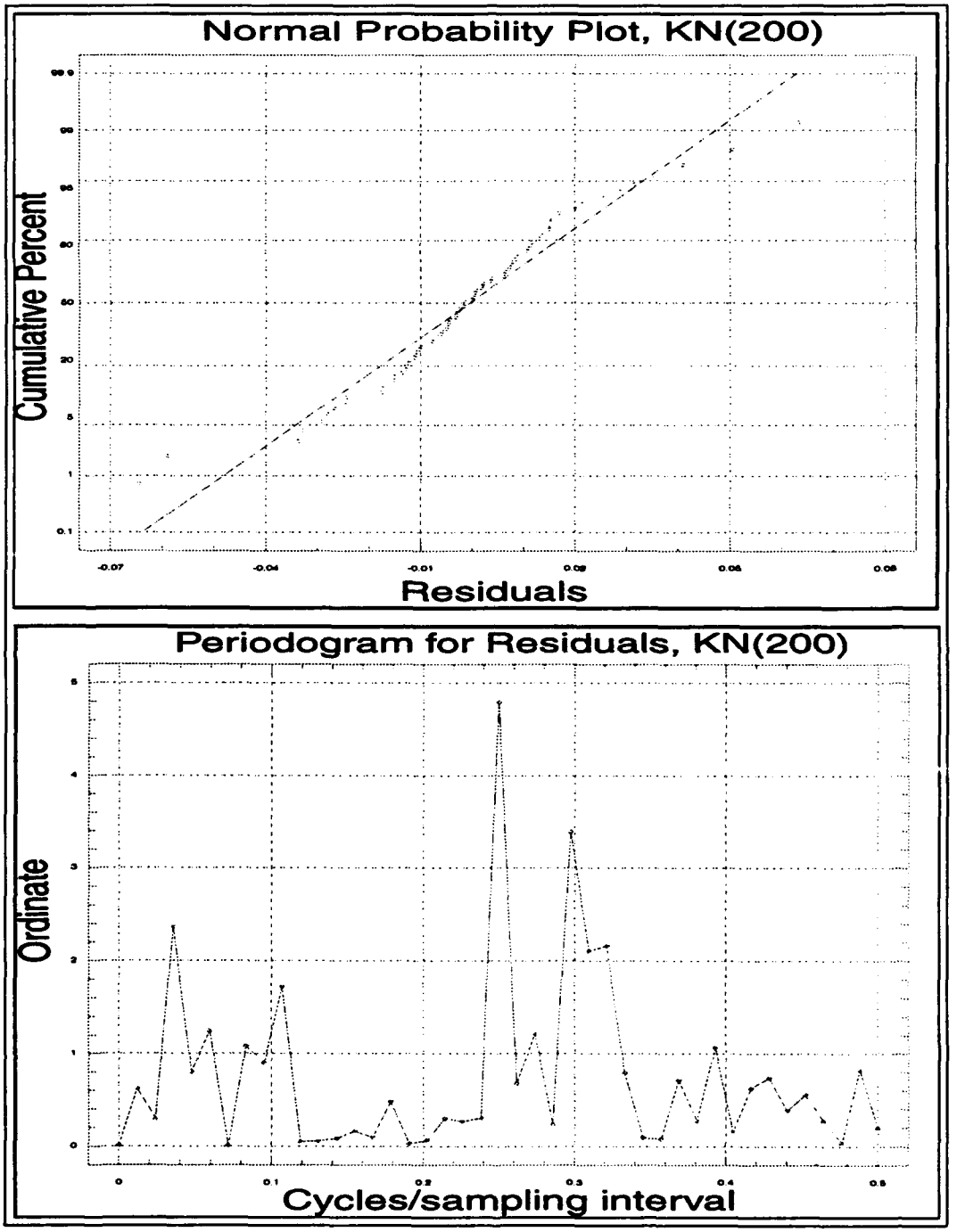

Figure F.118. Cumulative probability plots and periodograms for Model KN(2,0,0).

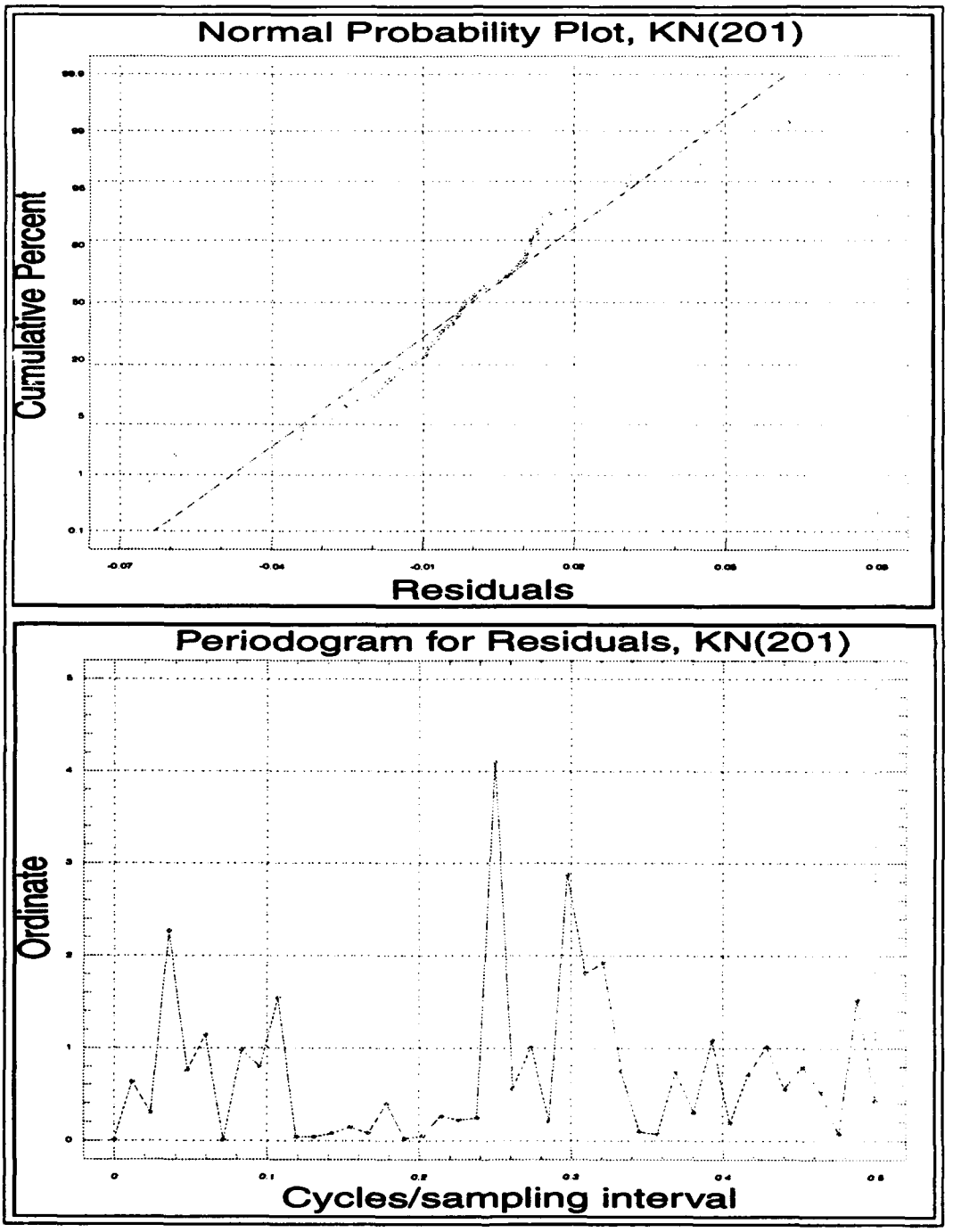

Figure F.119. Cumulative probability plots and periodograms for Model KN(2,0,1).

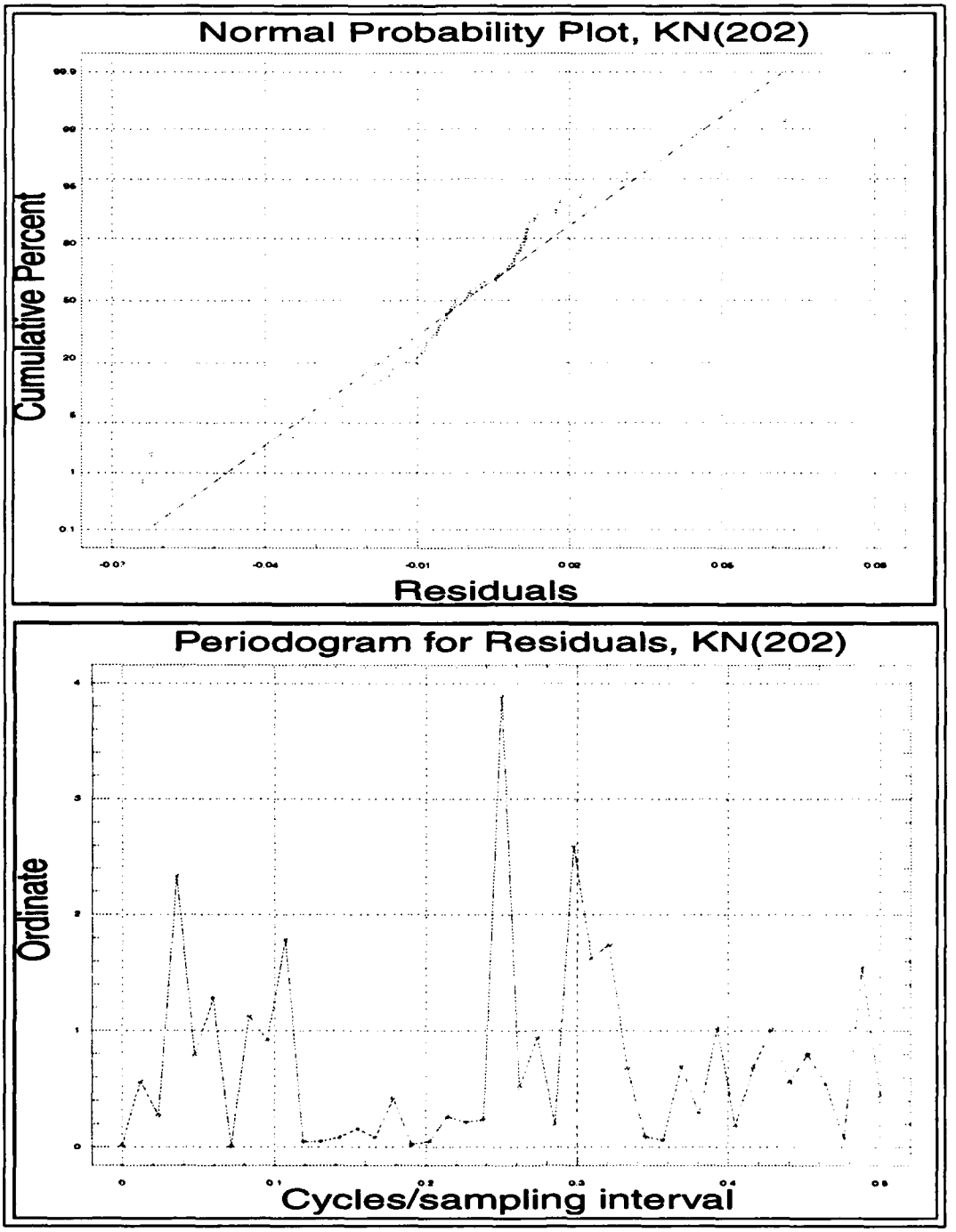

Figure F.120. Cumulative probability plots and periodograms for Model KN(2,0,2).

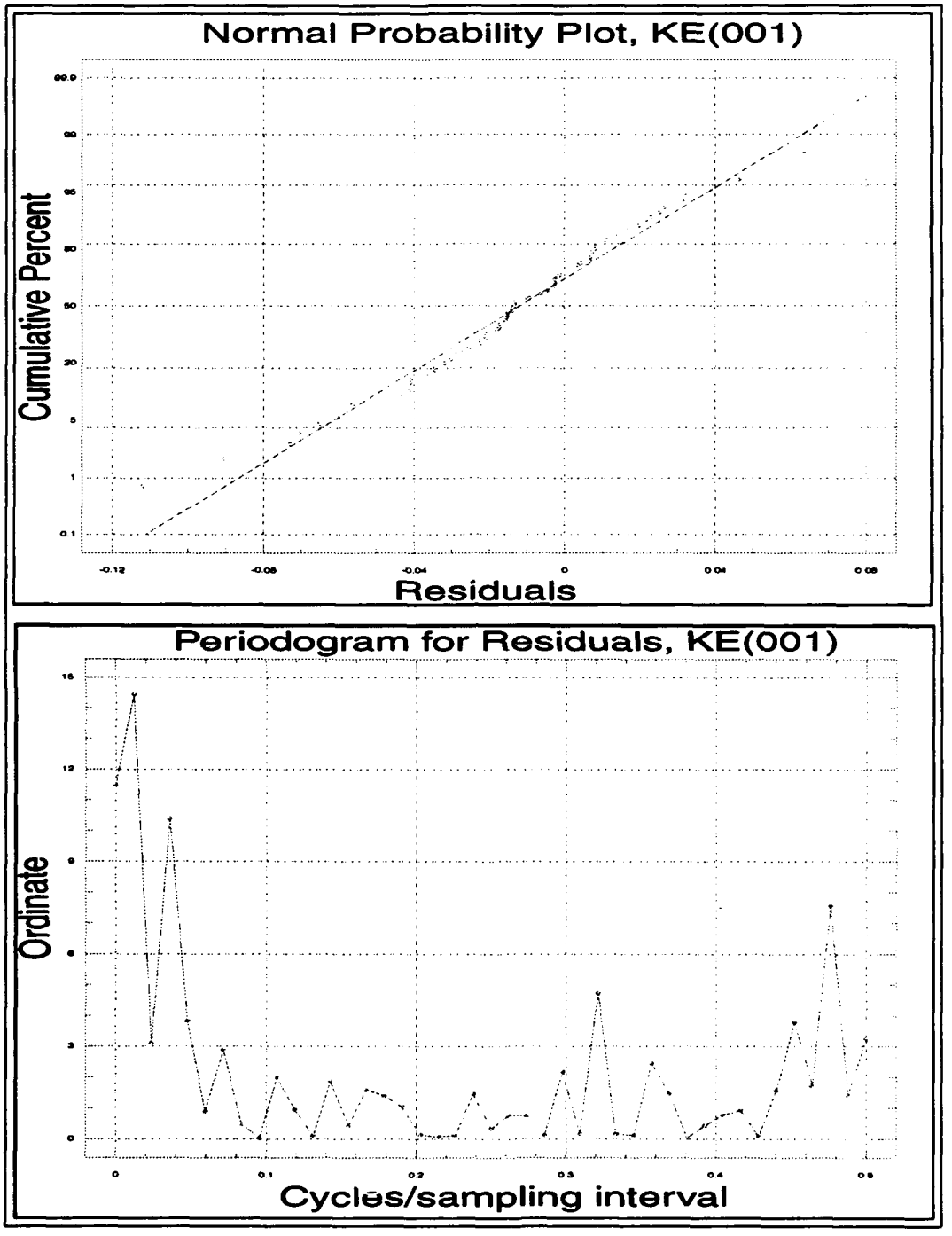

Figure F.121. Cumulative probability plots and periodograms for Model KE(0,0,1).

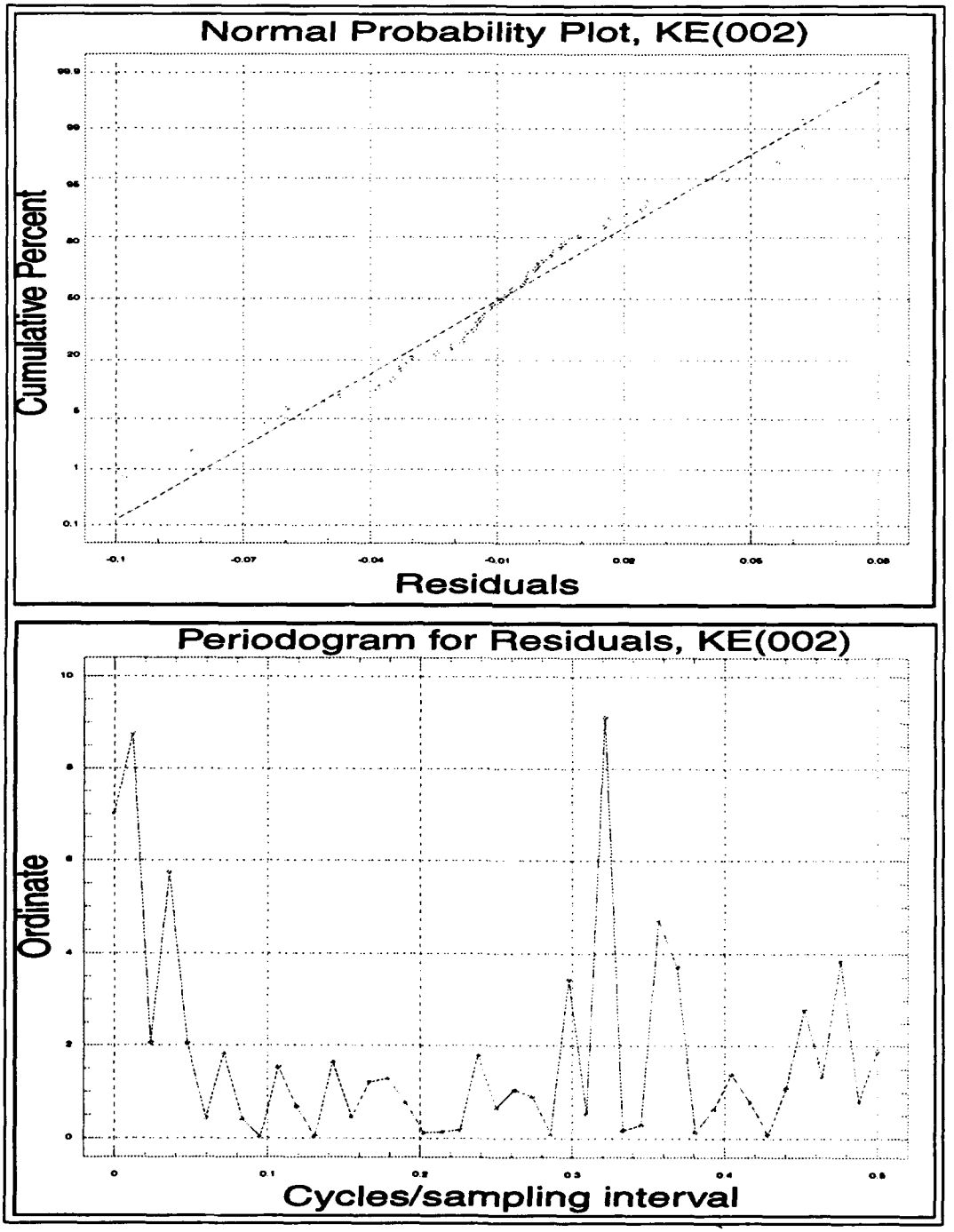

Figure F.122. Cumulative probability plots and periodograms for Model KE(0,0,2).

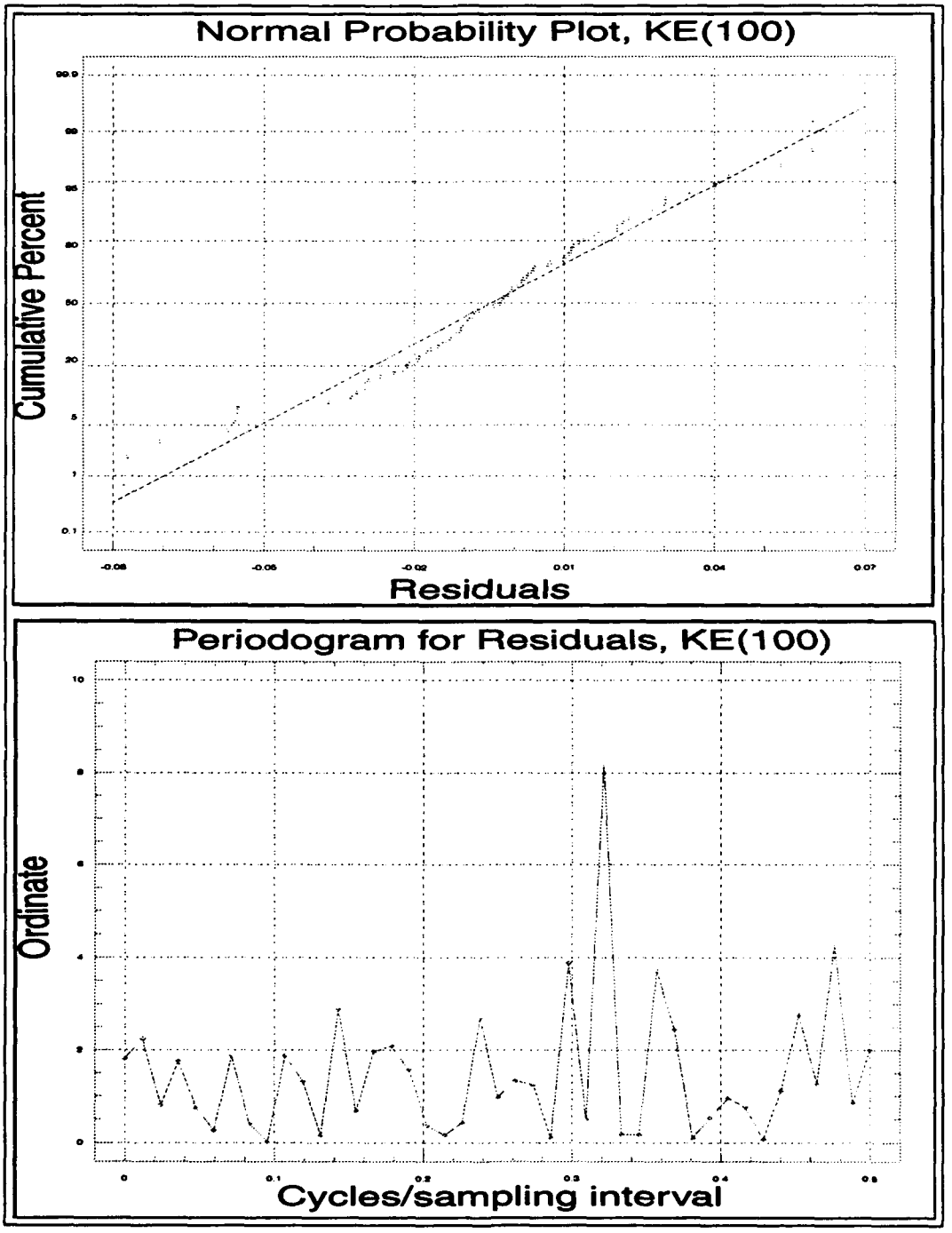

Figure F.123. Cumulative probability plots and periodograms for Model KE(1,0,0).

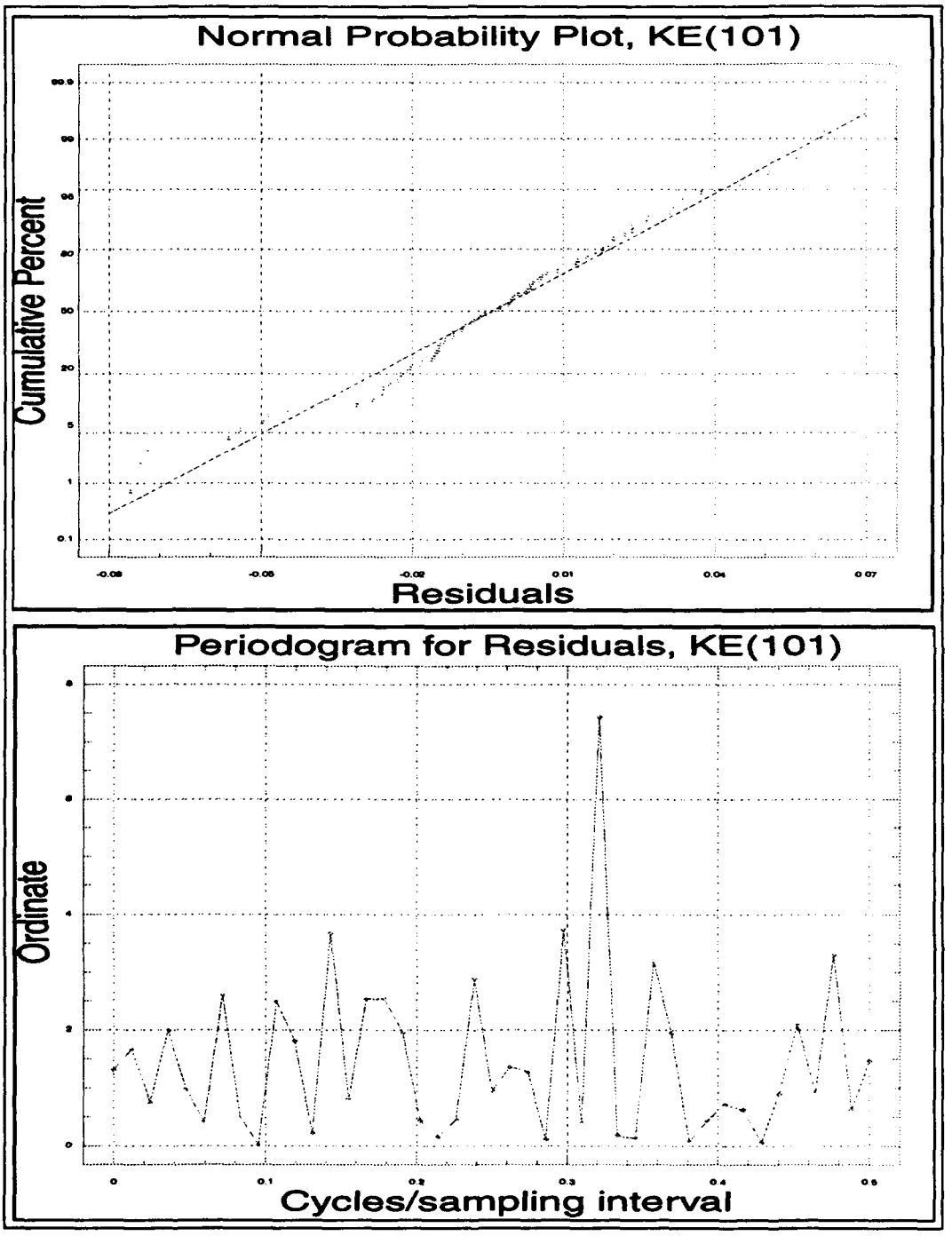

Figure F.124. Cumulative probability plots and periodograms for Model KE(1,0,1).

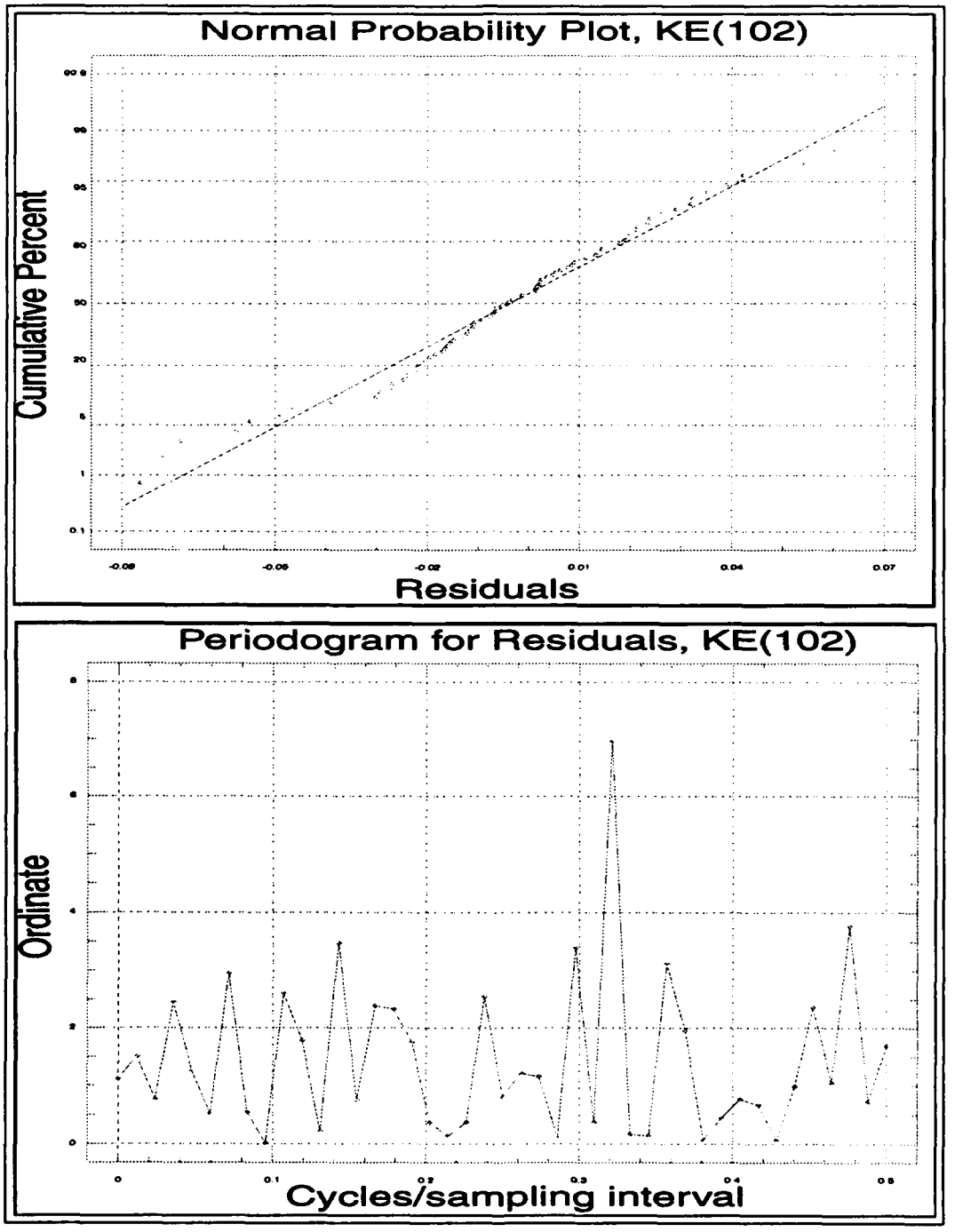

Figure F.125. Cumulative probability plots and periodograms for Model KE(1,0,2).

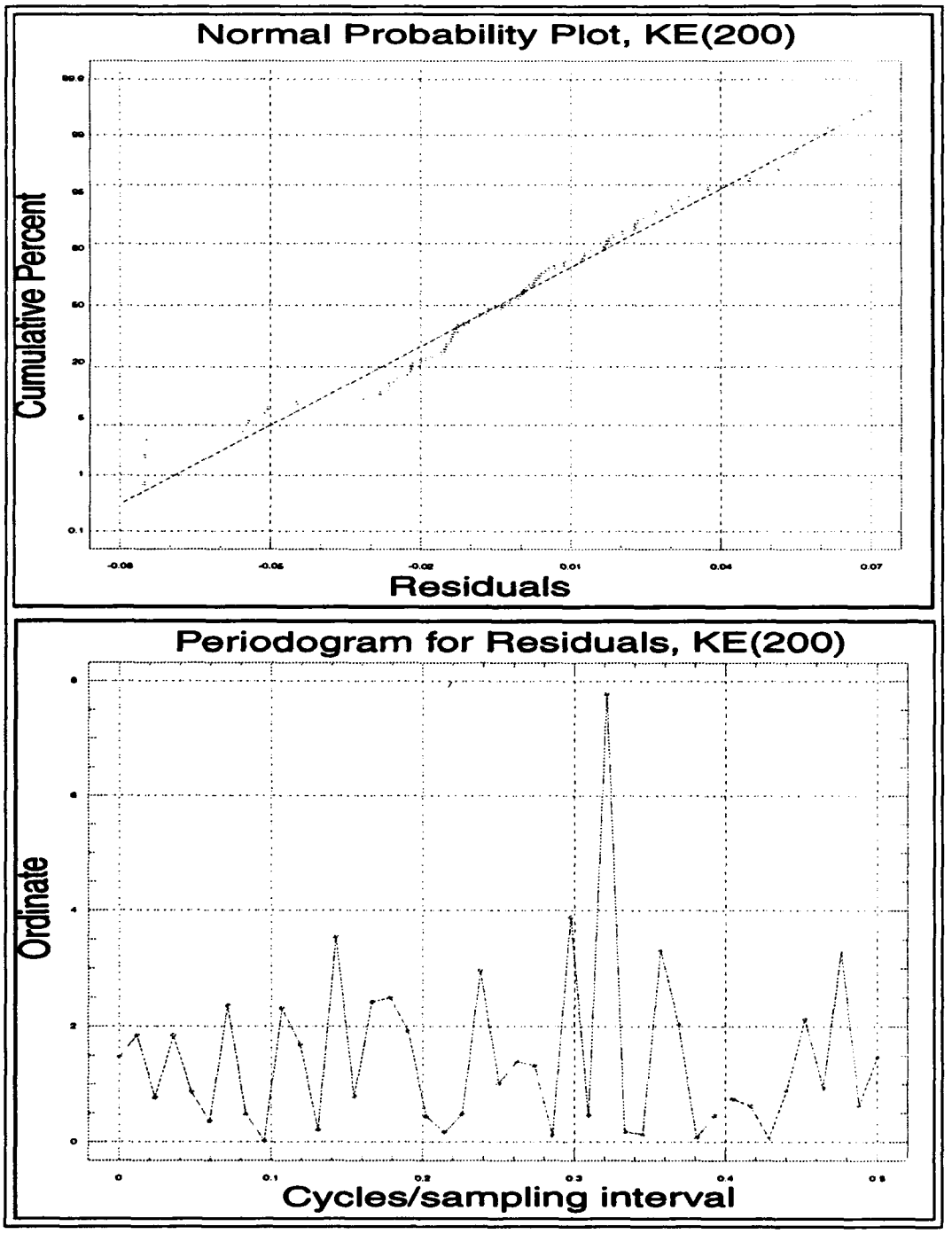

Figure F.126. Cumulative probability plots and periodograms for Model KE(2,0,0).

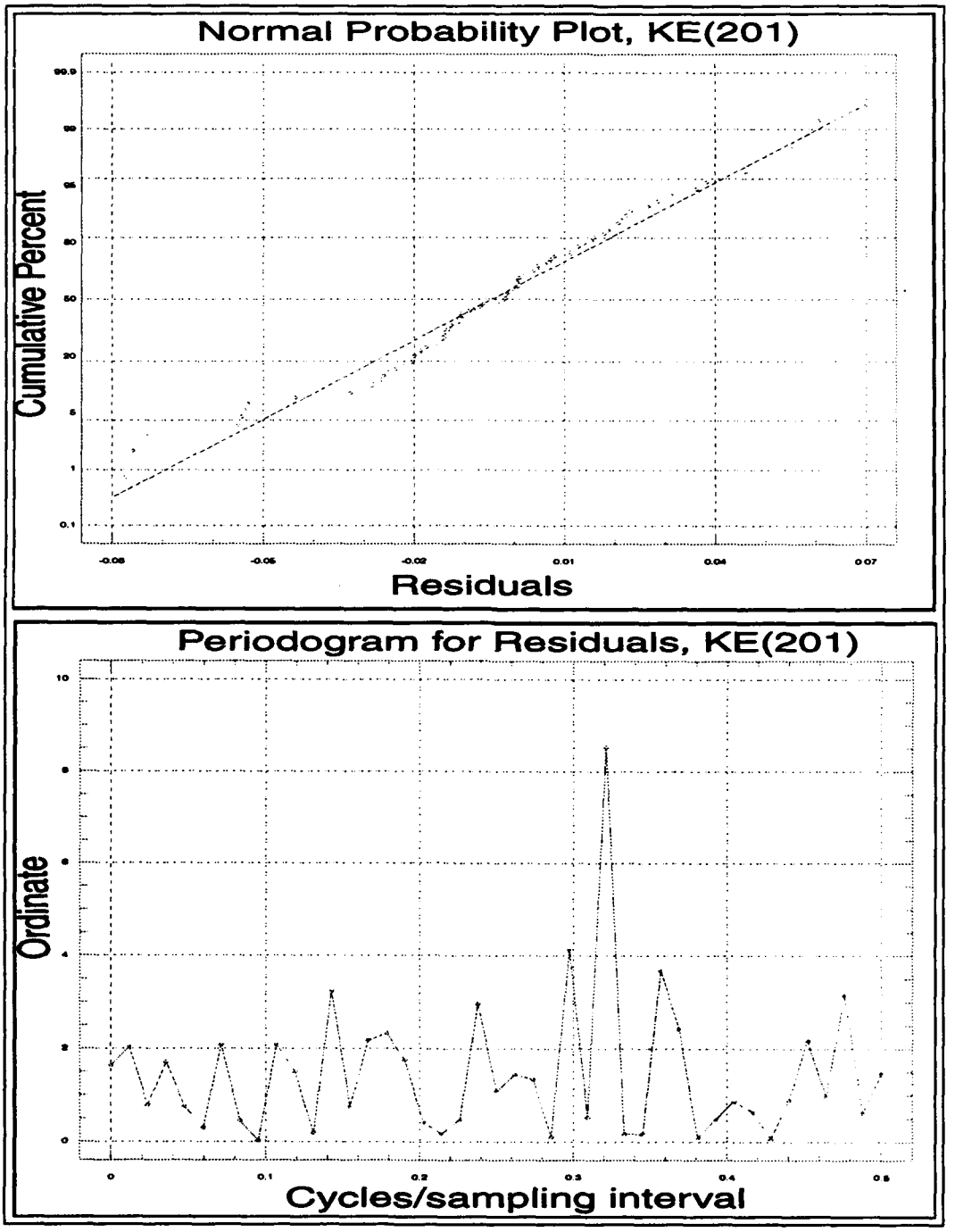

Figure F.127. Cumulative probability plots and periodograms for Model KE(2,0,1).

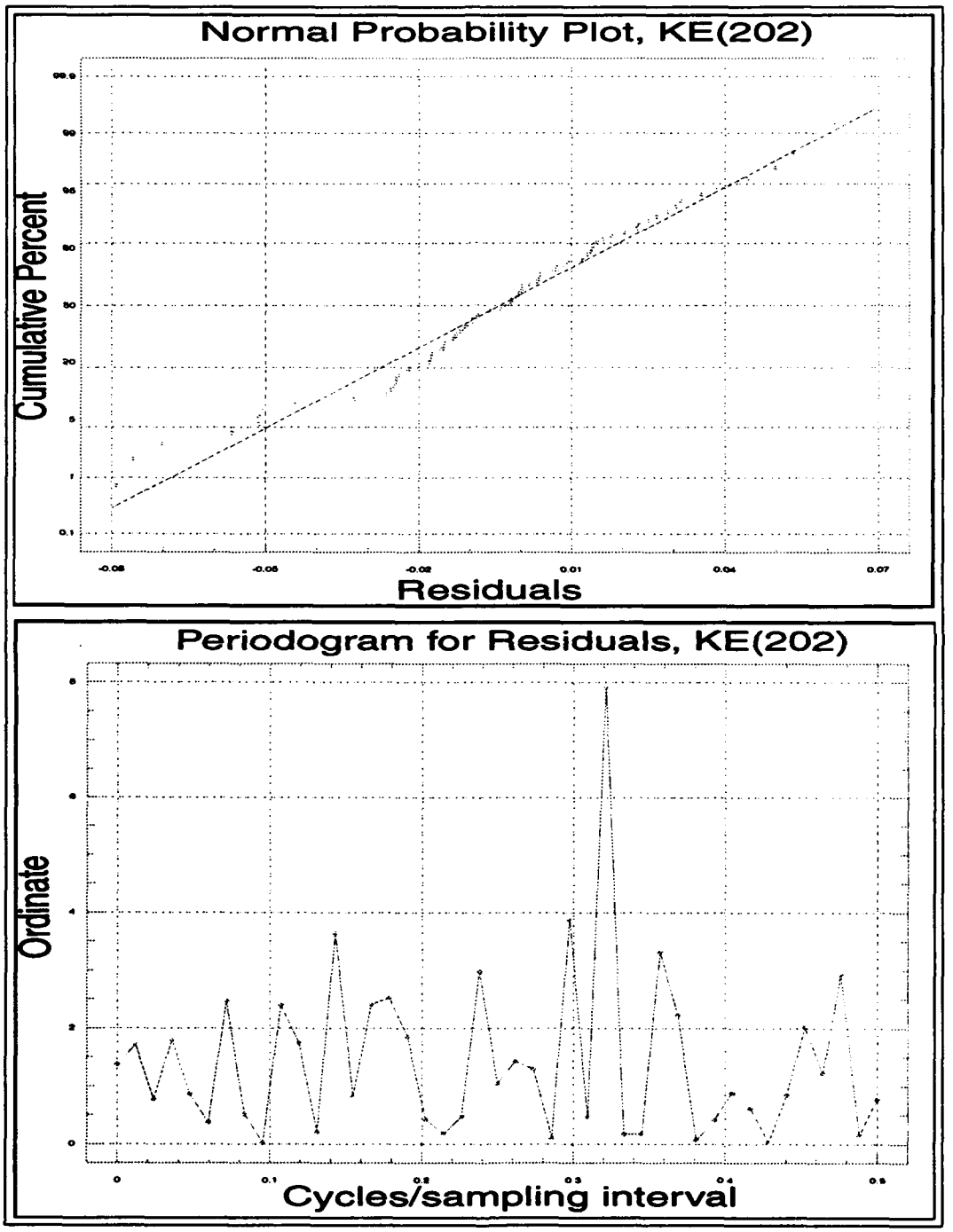

Figure F.128. Cumulative probability plots and periodograms for Model KE(2,0,2).

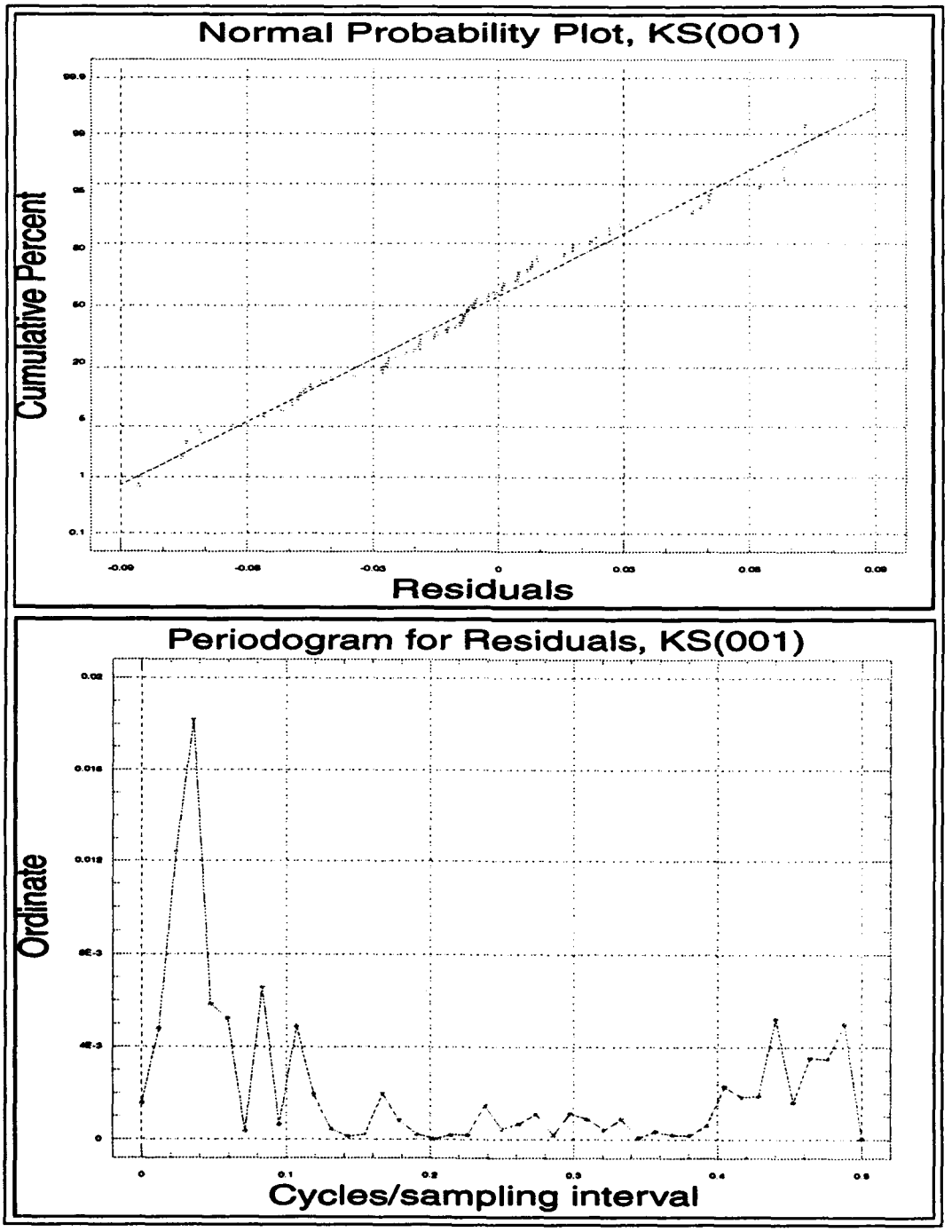

Figure F.129. Cumulative probability plots and periodograms for Model KS(0,0,1).

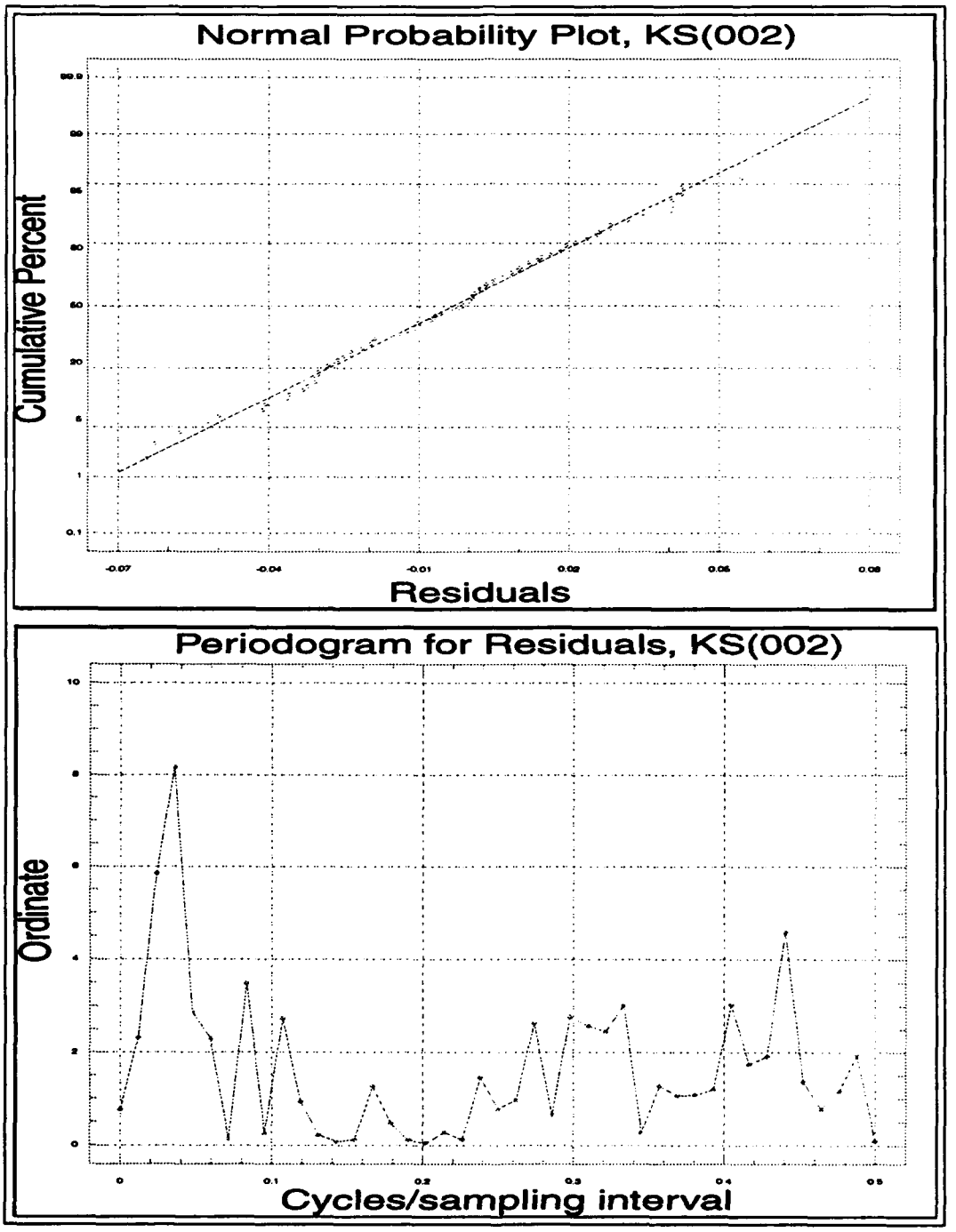

Figure F.130. Cumulative probability plots and periodograms for Model KS(0,0,2).

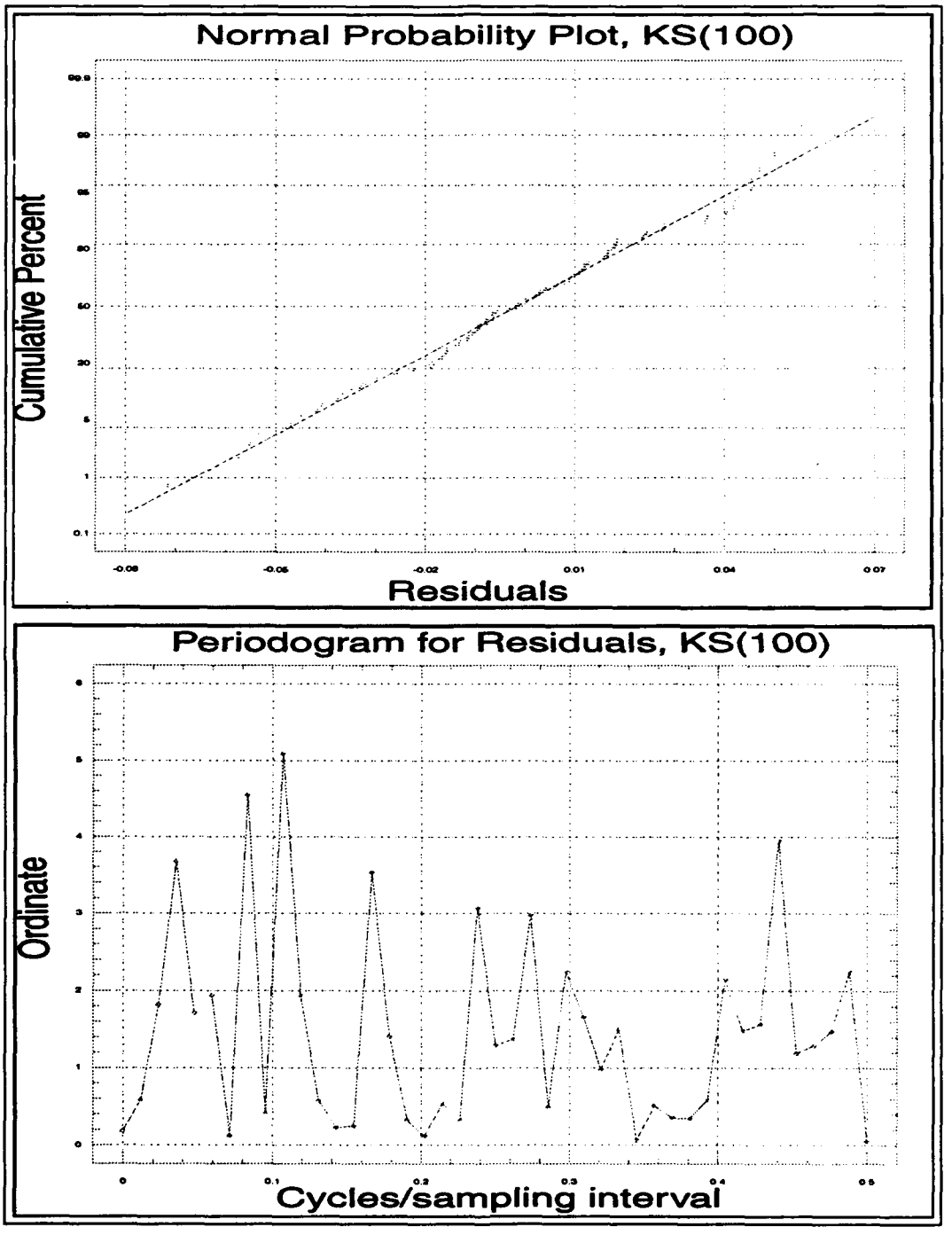

Figure F.131. Cumulative probability plots and periodograms for Model KS(1,0,0).

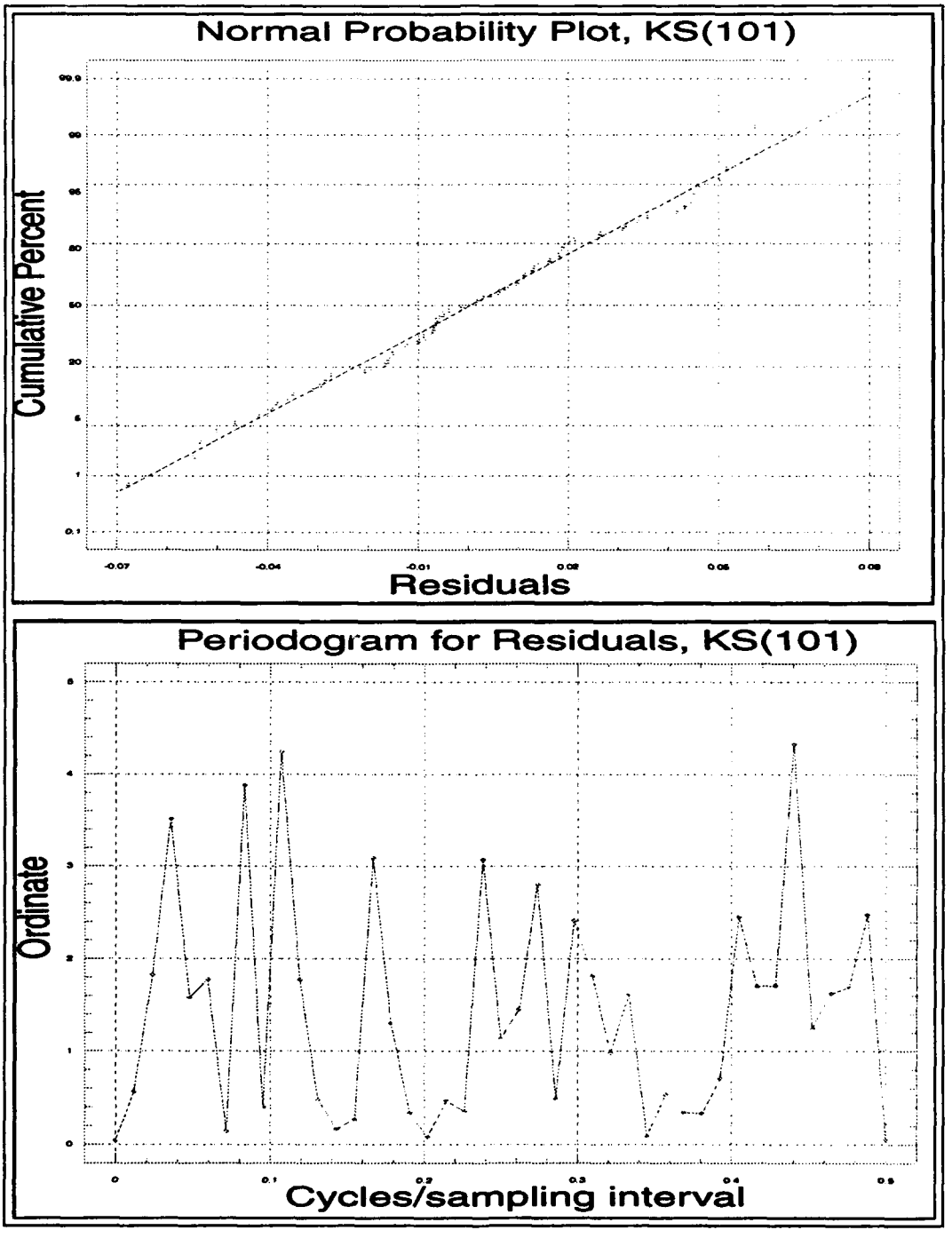

Figure F.132. Cumulative probability plots and periodograms for Model KS(1,0,1).

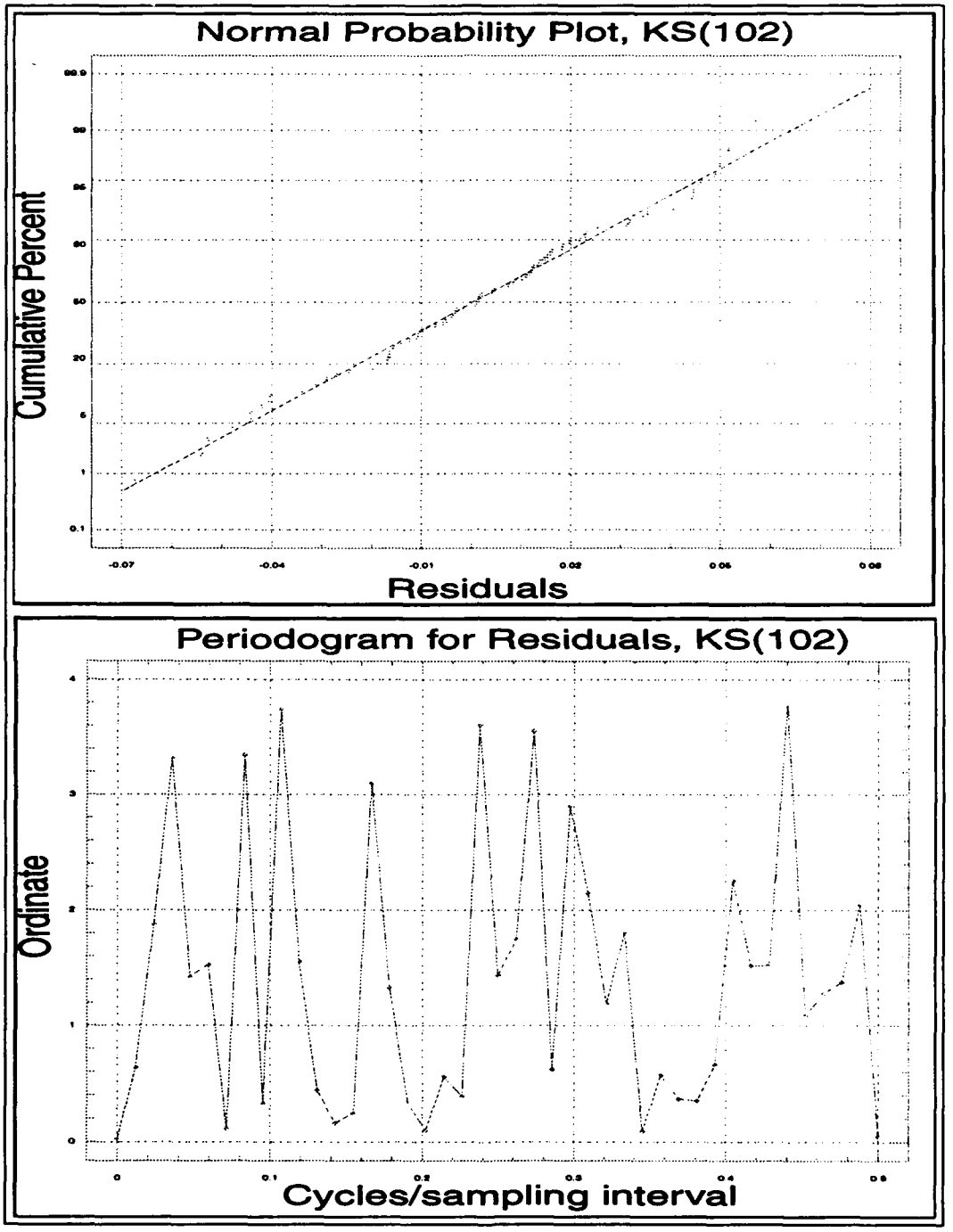

Figure F.133. Cumulative probability plots and periodograms for Model KS(1,0,2).

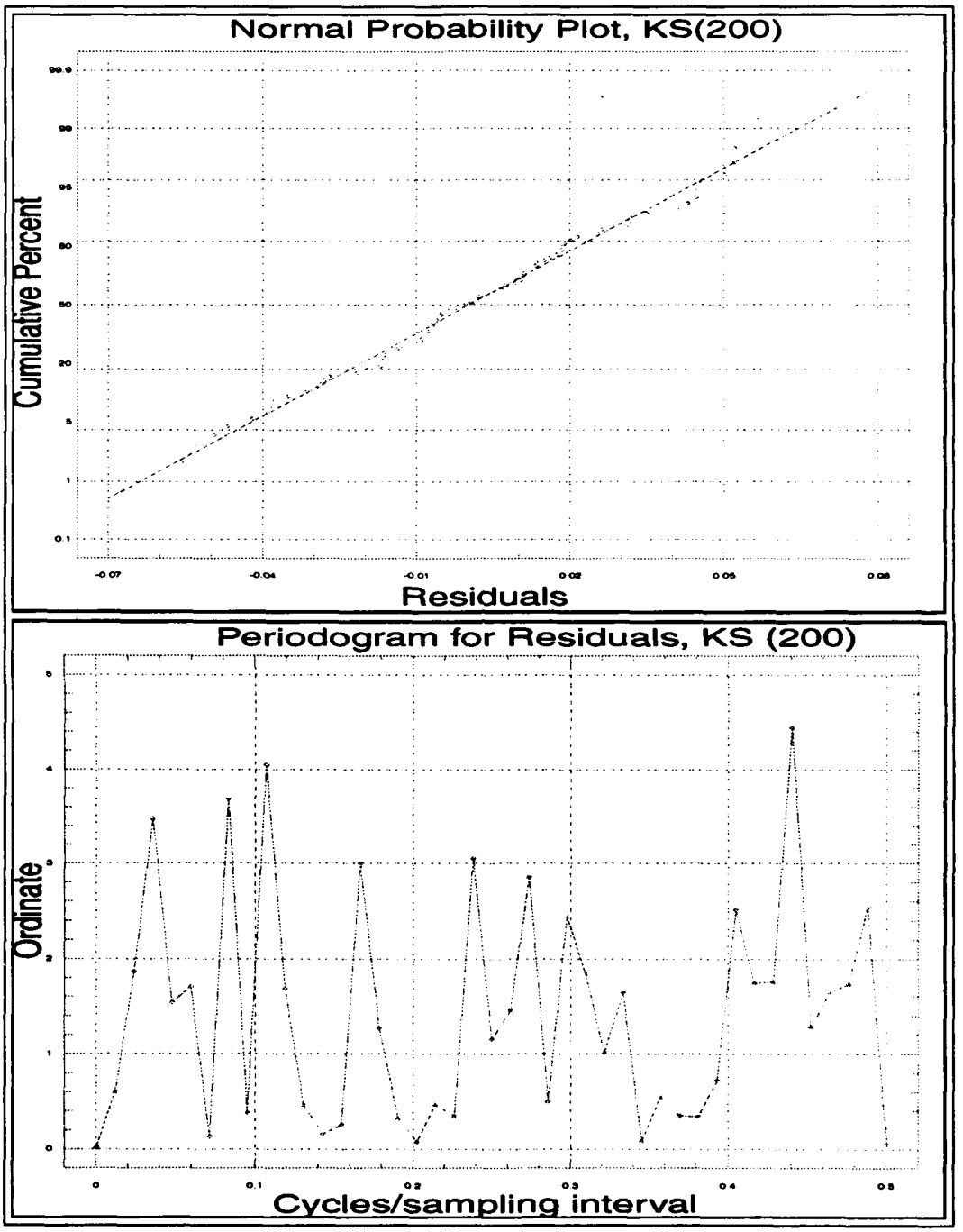

Figure F.134. Cumulative probability plots and periodograms for Model KS(2,0,0).

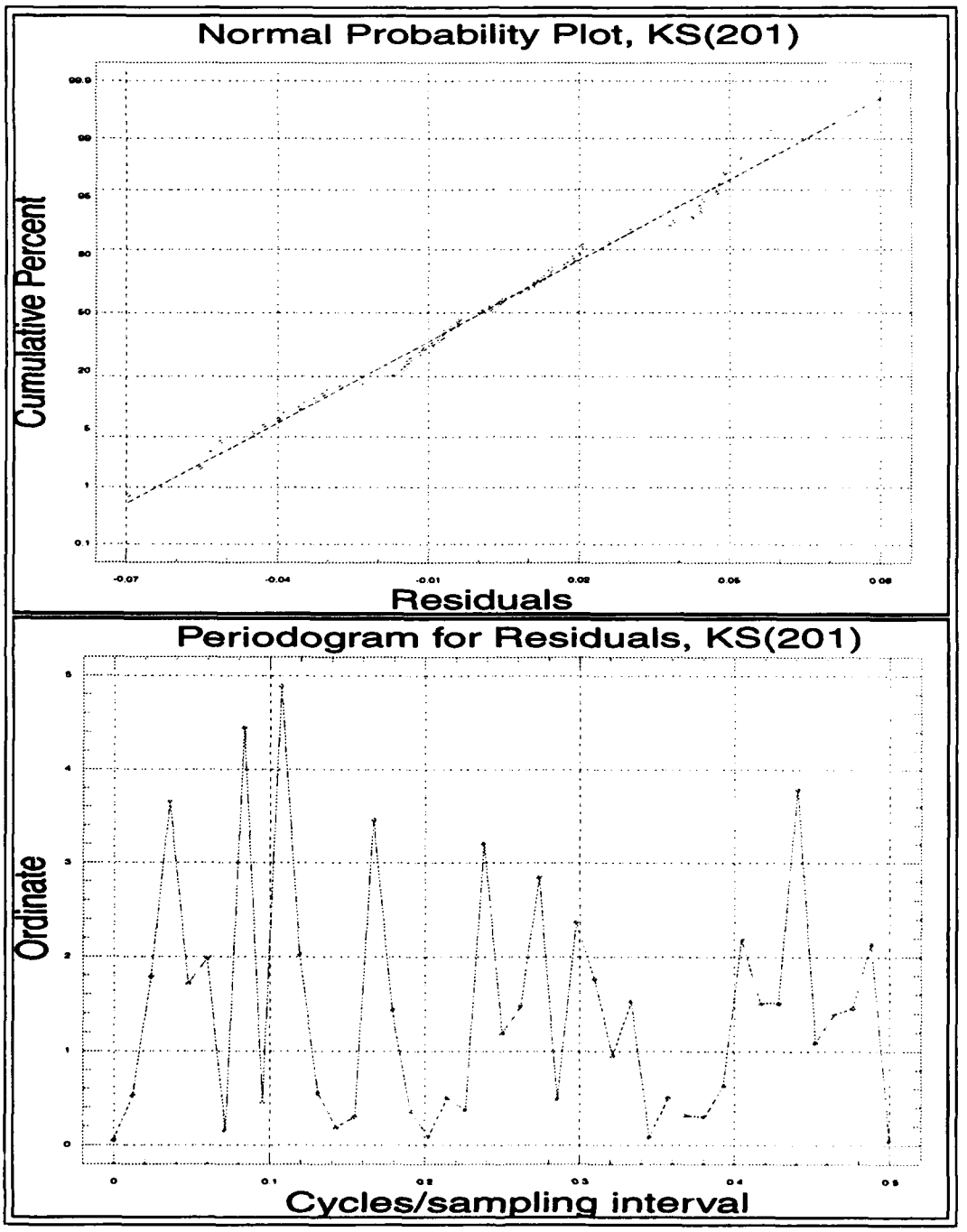

Figure F.135. Cumulative probability plots and periodograms for Model KS(2,0,1).

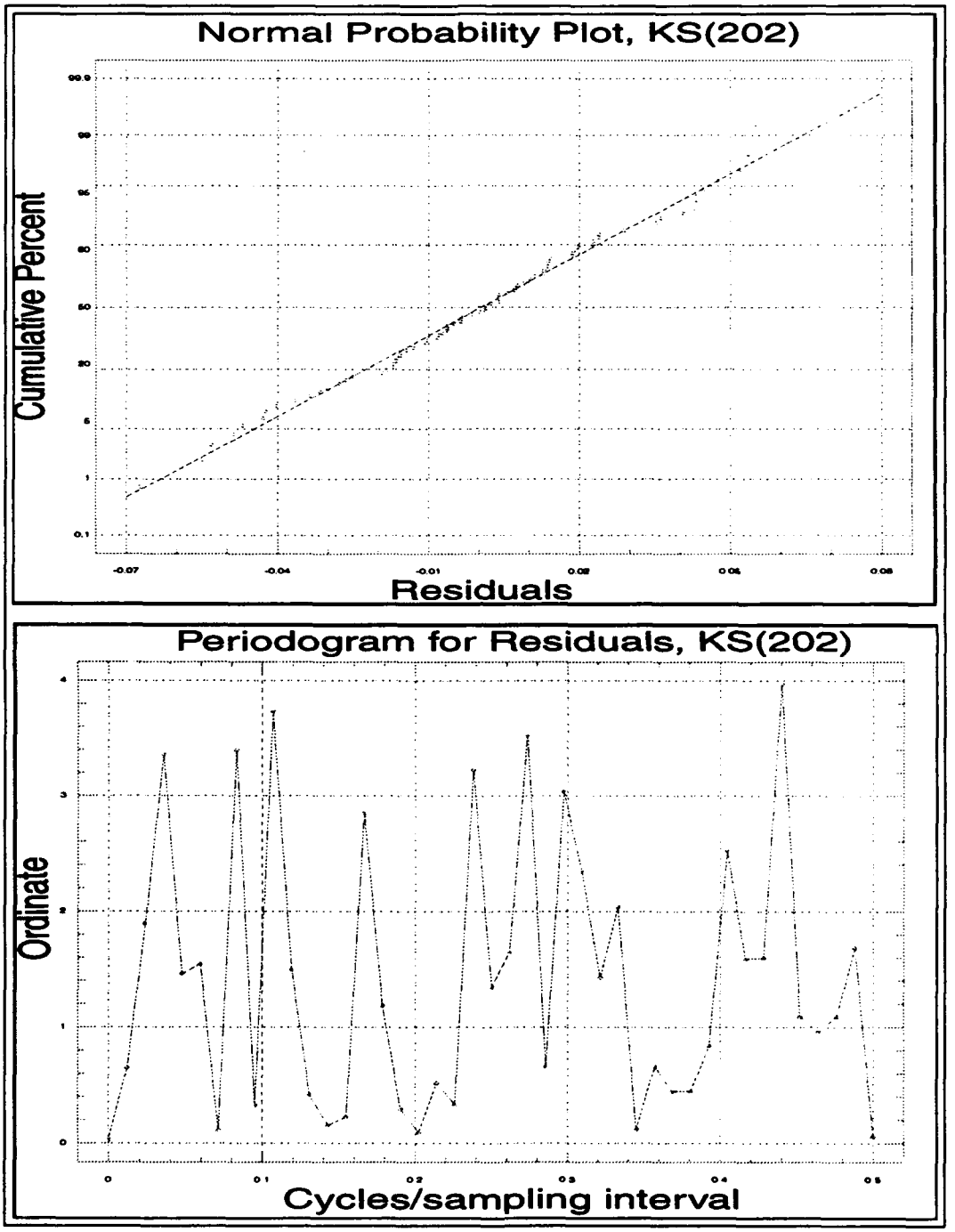

Figure F.136. Cumulative probability plots and periodograms for Model KS(2,0,2).

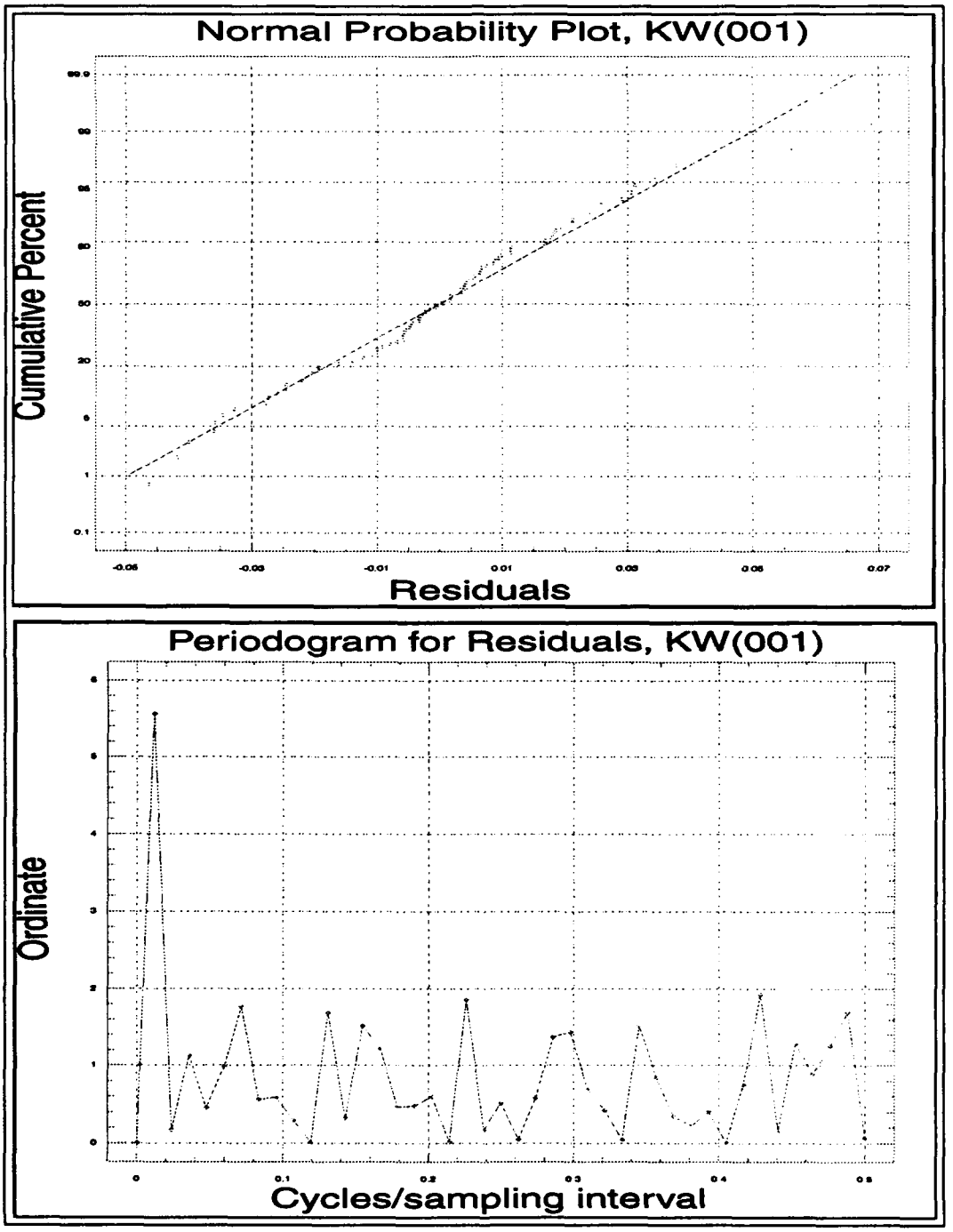

Figure F.137. Cumulative probability plots and periodograms for Model KW(0,0,1).

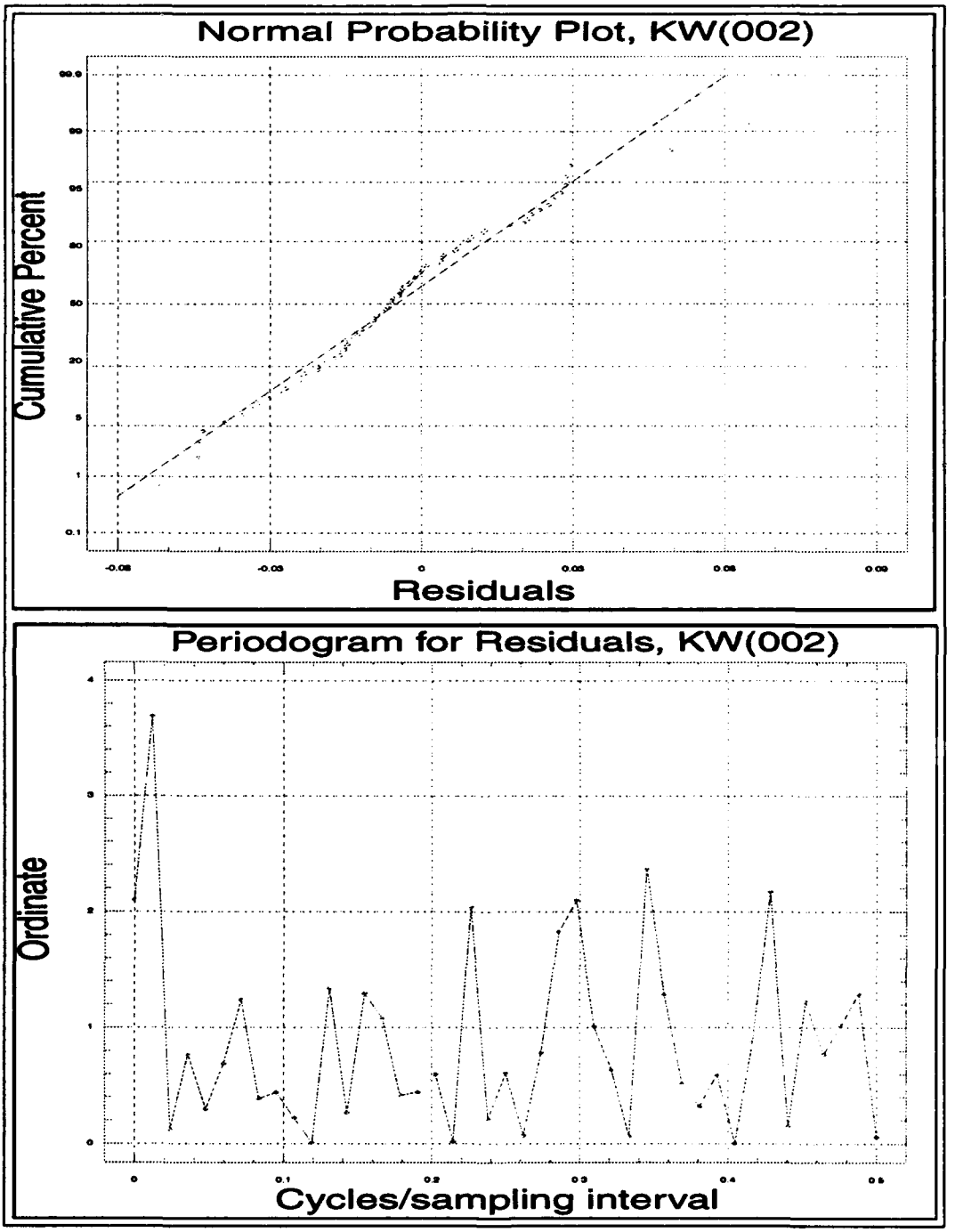

Figure F.138. Cumulative probability plots and periodograms for Model KW(0,0,2).

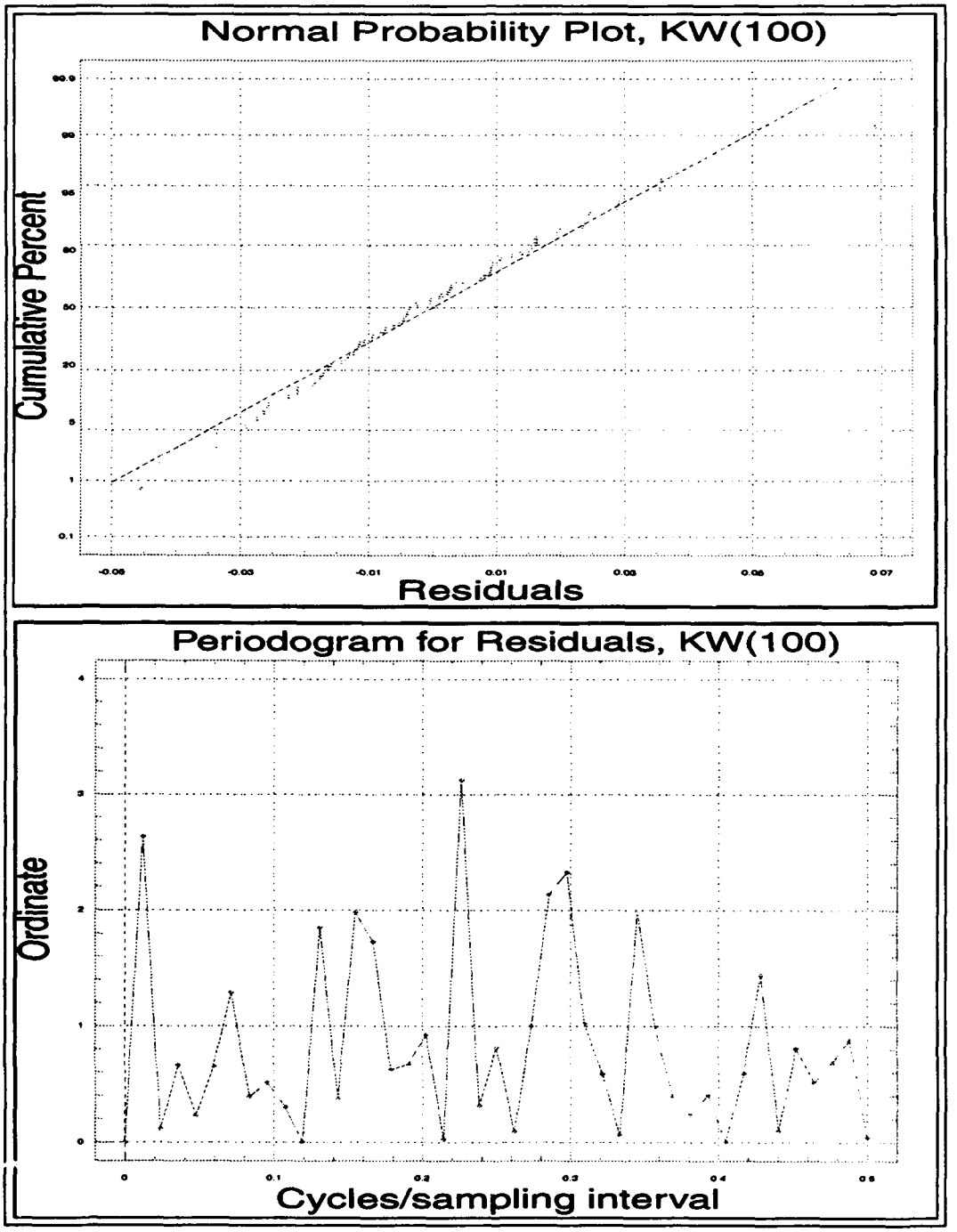

Figure F.139. Cumulative probability plots and periodograms for Model KW(1,0,0).

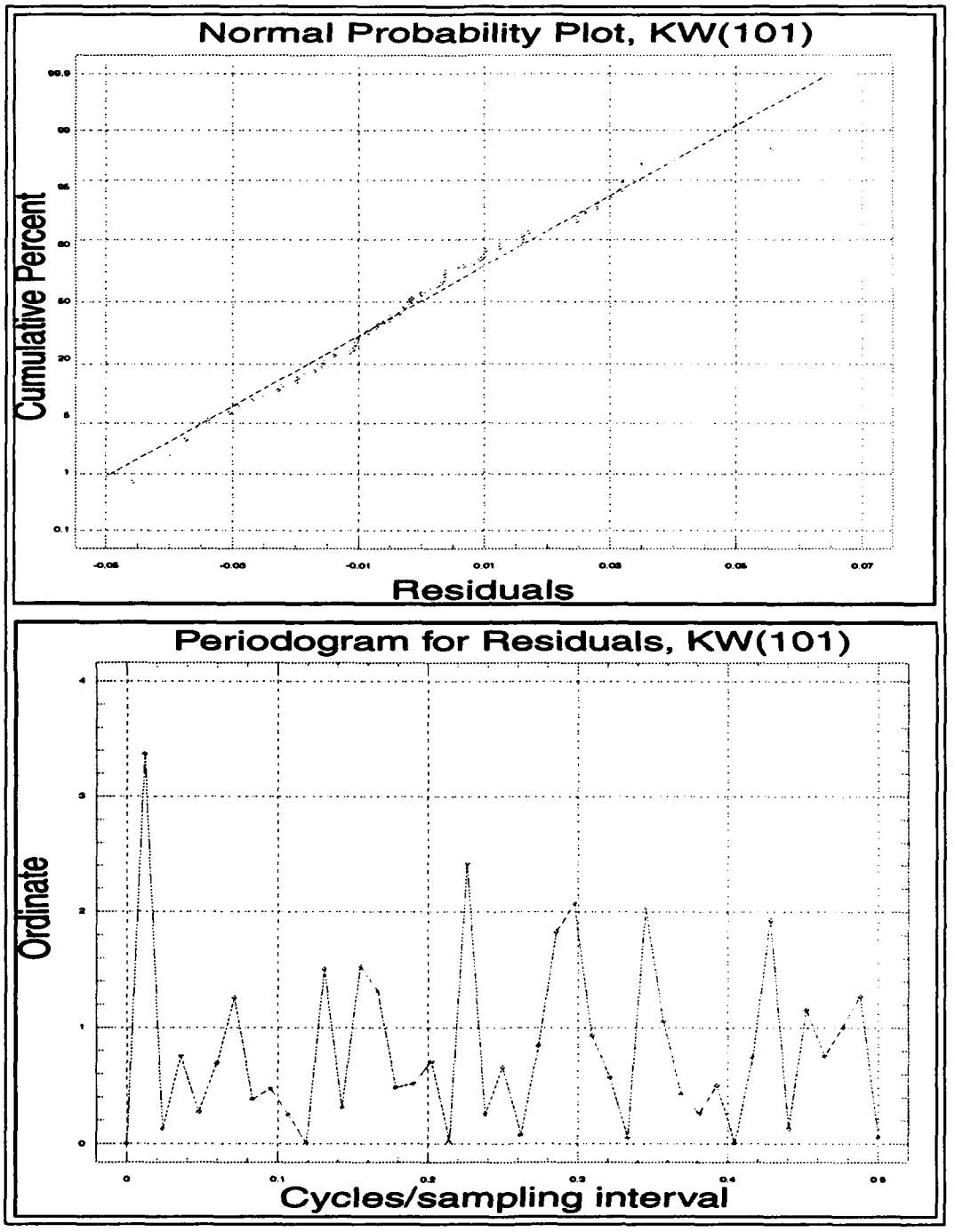

Figure F.140. Cumulative probability plots and periodograms for Model KW(1,0,1).

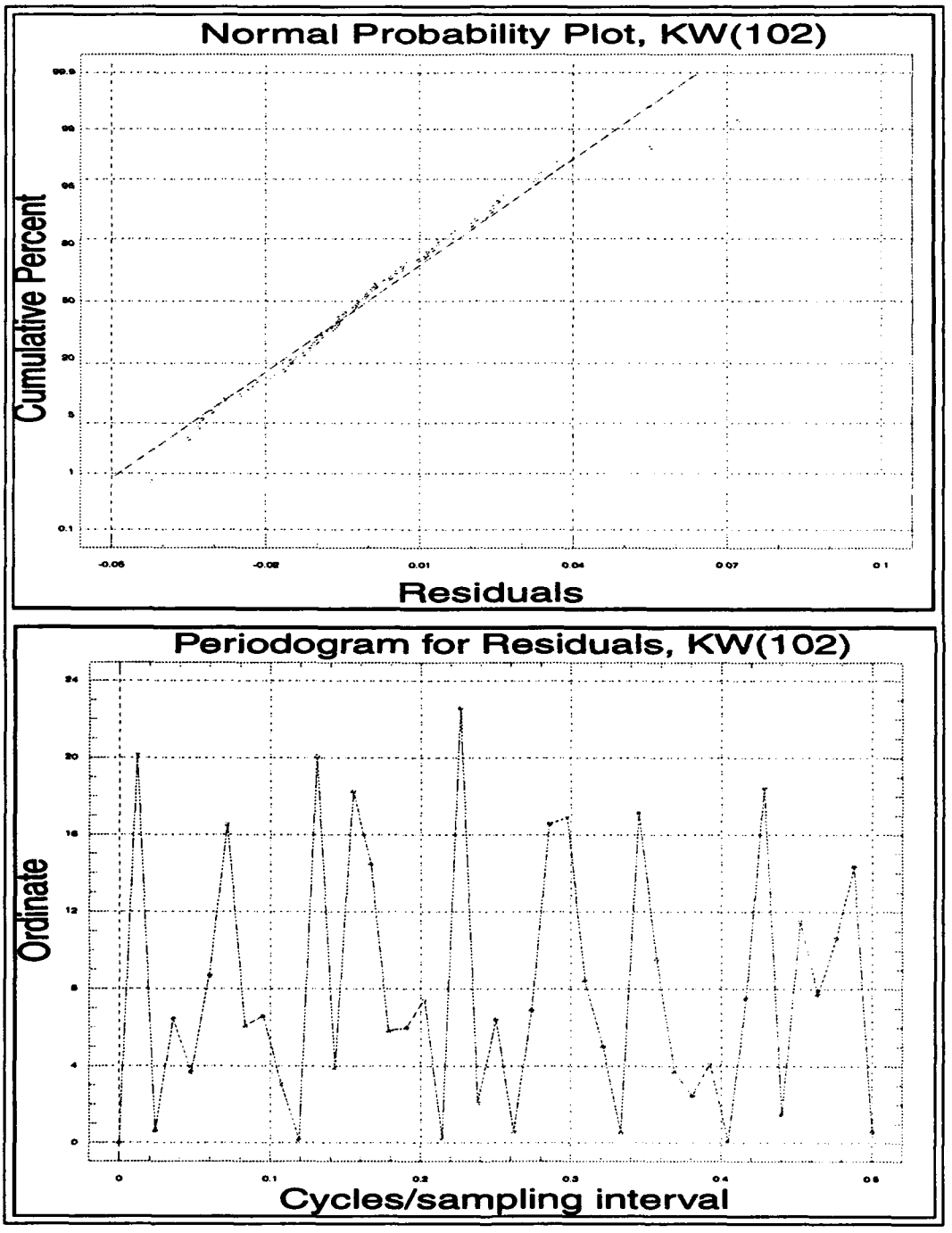

Figure F.141. Cumulative probability plots and periodograms for Model KW(1,0,2).

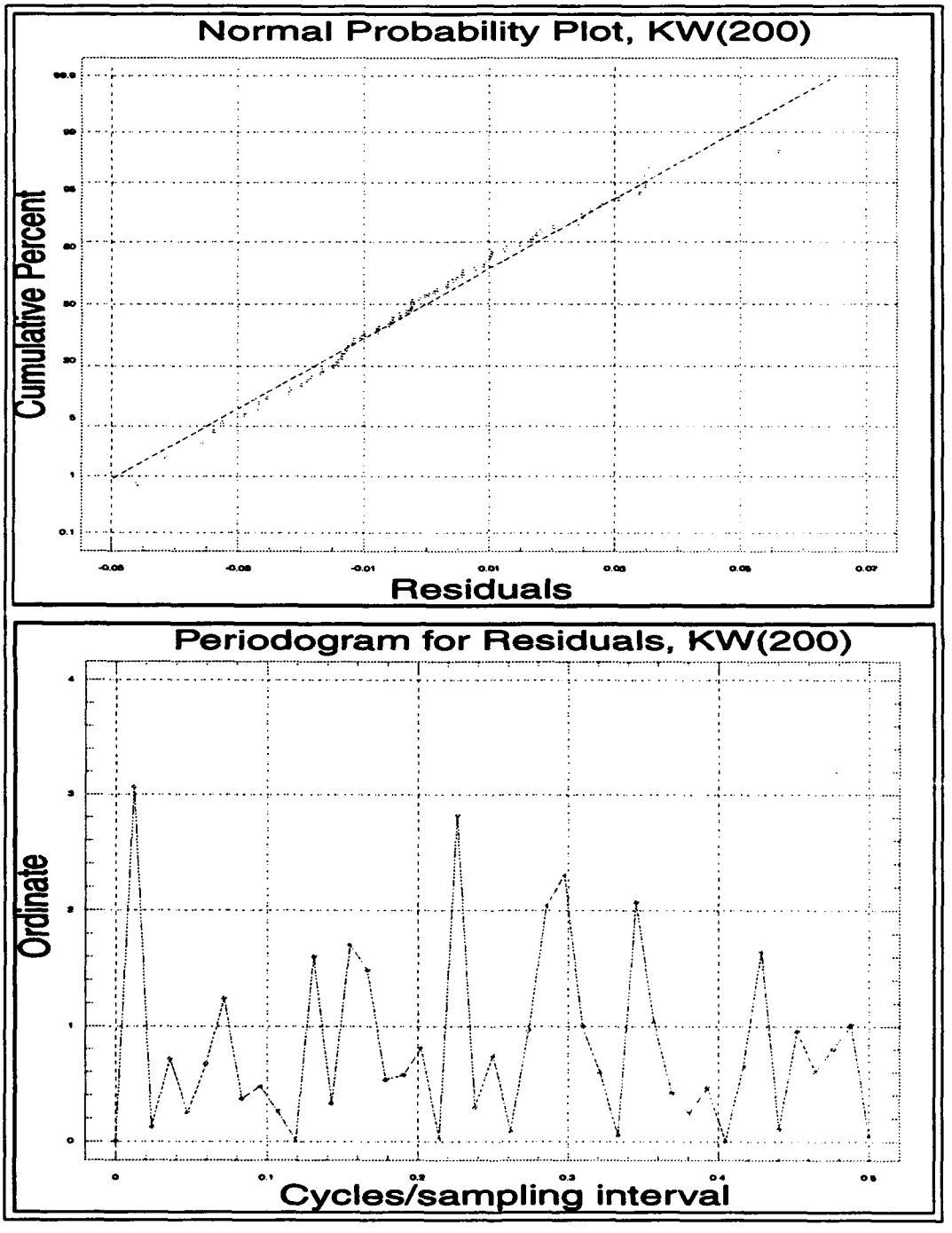

Figure F.142. Cumulative probability plots and periodograms for Model KW(2,0,0).

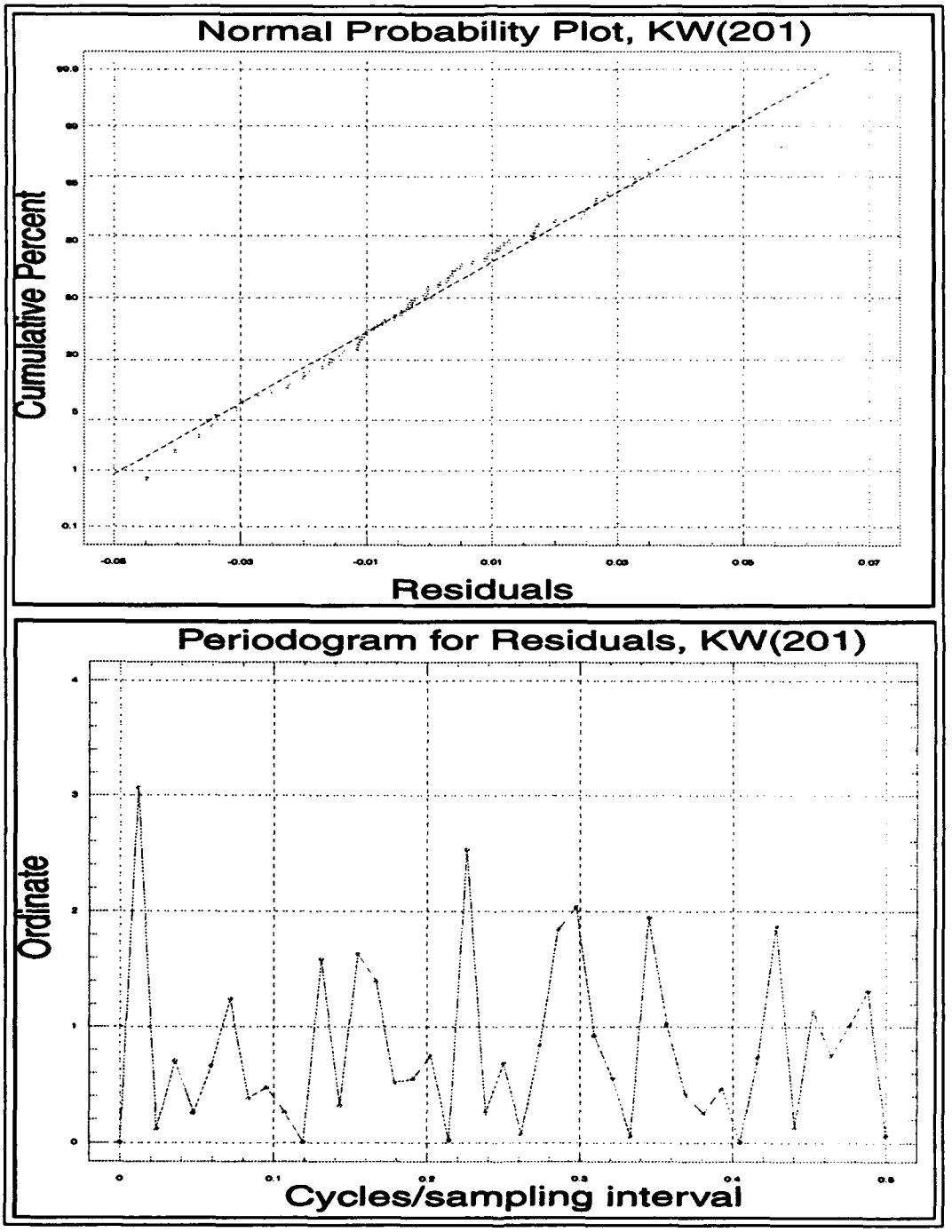

Figure F.143. Cumulative probability plots and periodograms for Model KW(2,0,1).

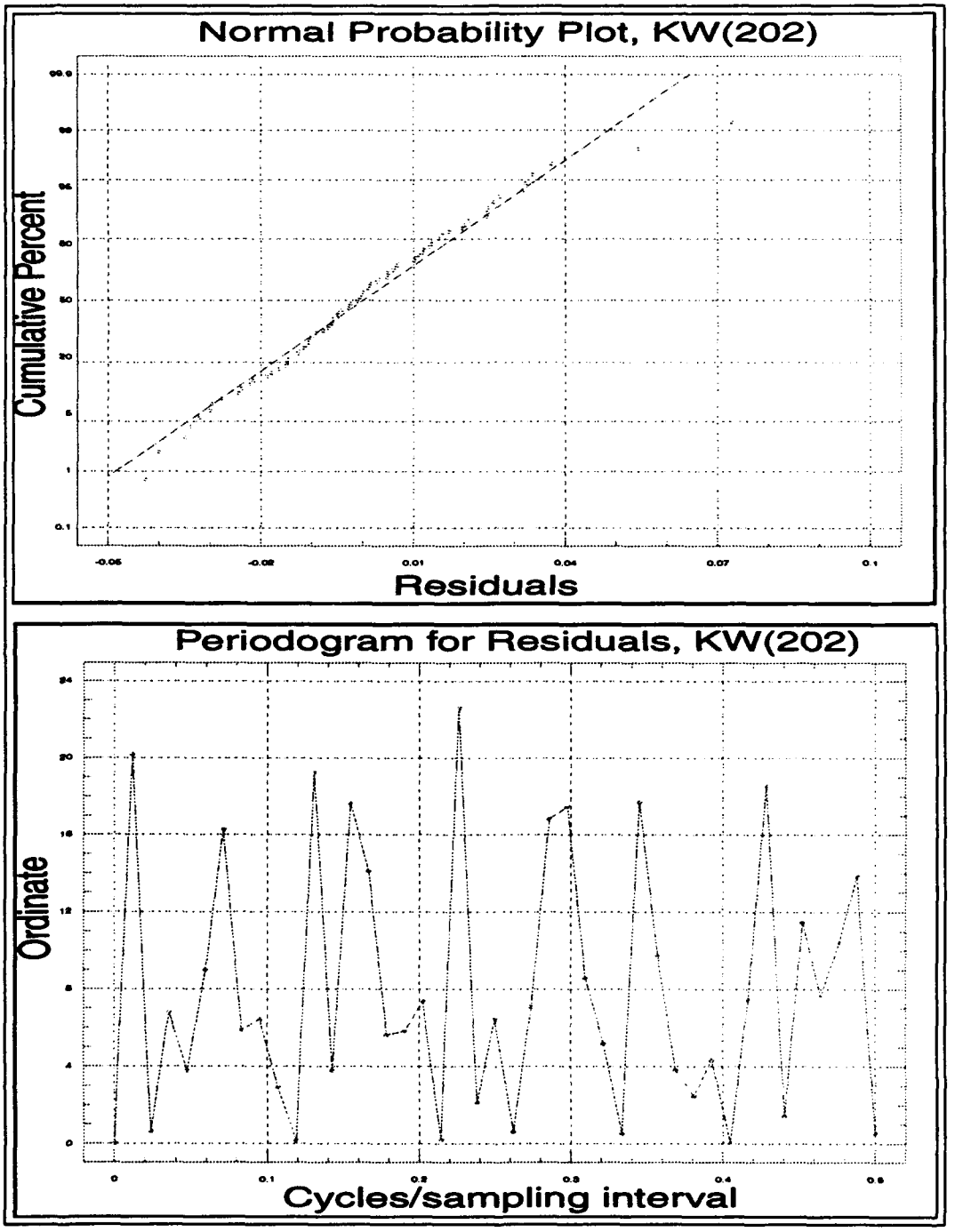

Figure F.144. Cumulative probability plots and periodograms for Model KW(2,0,2).

Appendix G. *Tests of Residuals*

The results of the five tests of residuals for the "best" model at each site are presented in this section. The order of presentation is Columbia data first, then Kirtland data; North, East, South, West. (Columbia East and South both have two entries.)

RUNS UP AND DOWN TEST		
Length	Observed	Expected
1	39	35.5000
2	12	15.3500
3 & up	6	5.4833

RUNS ABOVE AND BELOW THE MEAN TEST		
Length	Observed	Expected
1	16	21.7500
2	9	10.7500
3	4	5.31250
4 & up	8	5.18750

FREQUENCY TEST		
Interval	Observed	Expected
1	7	9.29050
2	7	9.46050
3	13	9.29900
4	7	9.71550
5	18	9.46900
6	11	9.71550
7	8	9.29900
8	7	9.46050
9	6	9.29050

PORTMANTEAU TEST		
dof	Q	Chi squared
1	0.03	2.71
2	0.07	4.61
3	0.11	6.25
4	0.14	7.78
5	0.21	8.24
6	0.24	10.64
7	0.39	12.02
8	0.44	13.36
9	0.51	14.68
10	0.90	15.99
11	0.98	17.27
12	1.08	18.55
13	3.43	19.81
14	3.73	21.06
15	4.00	22.31
16	4.26	23.54
17	4.57	24.77
18	4.83	25.99
19	5.14	27.20
20	5.40	28.41

AIC= -473.29

Figure G.1. Results of the tests of residuals for Model CN(100).

RUNS UP AND DOWN TEST		
Length	Observed	Expected
1	34	35.5000
2	16	15.3500
3 & up	6	5.4833

RUNS ABOVE AND BELOW THE MEAN TEST		
Length	Observed	Expected
1	21	21.7500
2	15	10.7500
3	4	5.31250
4 & up	4	5.18750

FREQUENCY TEST		
Interval	Observed	Expected
1	8	9.29050
2	9	9.46050
3	9	9.29900
4	11	9.71550
5	12	9.46900
6	9	9.71550
7	12	9.29900
8	5	9.46050
9	10	9.29050

PORTMANTEAU TEST		
dof	Q	Chi squared
1	0.09	2.71
2	0.14	4.61
3	0.20	6.25
4	0.40	7.78
5	0.49	8.24
6	0.79	10.64
7	0.88	12.02
8	1.18	13.36
9	2.25	14.68
10	2.98	15.99
11	3.21	17.27
12	3.67	18.55
13	4.03	19.81
14	4.46	21.06
15	4.71	22.31
16	4.99	23.54
17	5.34	24.77
18	5.68	25.99
19	5.96	27.20
20	6.24	28.41

AIC= -396.33

Figure G.2. Results of the tests of residuals for Model CE(202).

RUNS UP AND DOWN TEST

Length	Observed	Expected
1	49	35.5000
2	14	15.3500
3	2	5.4833

RUNS ABOVE AND BELOW THE MEAN TEST

Length	Observed	Expected
1	30	21.7500
2	8	10.7500
3	2	5.3125
4	6	5.1875

FREQUENCY TEST

Length	Observed	Expected
1	9	9.29050
2	6	9.46050
3	11	9.29900
4	9	9.715500
5	15	9.46900
6	9	9.715500
7	9	9.29900
8	9	9.46050
9	8	9.29050

PORTMANTEAU TEST

dof	Q	Chi squared
1	0.05	2.71
2	0.07	4.61
3	0.33	6.25
4	0.41	7.78
5	0.58	8.24
6	0.73	10.64
7	1.22	12.02
8	1.37	13.36
9	1.94	14.68
10	2.13	15.99
11	2.51	17.27
12	3.56	18.55
13	4.87	19.81
14	5.27	21.06
15	5.93	22.31
16	6.33	23.54
17	7.10	24.77
18	7.52	25.99
19	8.03	27.20
20	8.53	28.41

AIC= -395.57

Figure G.3. Results of the tests of residuals for Model CE(100).

RUNS UP AND DOWN TEST

Length	Observed	Expected
1	28	35.5000
2	15	15.3500
3 & up	5	5.4833

RUNS ABOVE AND BELOW THE MEAN TEST

Length	Observed	Expected
1	17	21.7500
2	8	10.7500
3	4	5.31250
4 & up	9	5.18750

FREQUENCY TEST

Interval	Observed	Expected
1	12	9.29050
2	4	9.46050
3	6	9.29900
4	10	9.71550
5	14	9.46900
6	6	9.71550
7	14	9.29900
8	11	9.46050
9	7	9.29050

PORTMANTEAU TEST

dof	Q	Chi squared
1	0.03	2.71
2	0.08	4.61
3	0.15	6.25
4	0.19	7.78
5	0.27	8.24
6	0.34	10.64
7	0.39	12.02
8	0.80	13.36
9	0.88	14.68
10	1.00	15.99
11	1.26	17.27
12	1.79	18.55
13	2.07	19.81
14	2.21	21.06
15	2.38	22.31
16	2.53	23.54
17	2.69	24.77
18	2.93	25.99
19	3.29	27.20
20	3.98	28.41

AIC= -423.29

Figure G.4. Results of the tests of residuals for Model CS(200).

RUNS UP AND DOWN TEST

Length	Observed	Expected
1	46	35.5000
2	16	15.3500
3	2	5.4833

RUNS ABOVE AND BELOW THE MEAN TEST

Length	Observed	Expected
1	26	21.7500
2	9	10.7500
3	3	5.3125
4	7	5.1875

FREQUENCY TEST

Length	Observed	Expected
1	11	9.29050
2	3	9.46050
3	8	9.29900
4	12	9.715500
5	9	9.46900
6	12	9.715500
7	11	9.29900
8	9	9.46050
9	9	9.29050

PORTMANTEAU TEST

dof	Q	Chi squared
1	0.09	2.71
2	0.14	4.61
3	0.20	6.25
4	0.33	7.78
5	0.40	8.24
6	0.51	10.64
7	0.65	12.02
8	0.79	13.36
9	1.22	14.68
10	1.38	15.99
11	1.52	17.27
12	1.74	18.55
13	2.16	19.81
14	2.44	21.06
15	2.61	22.31
16	2.86	23.54
17	3.04	24.77
18	3.24	25.99
19	3.50	27.20
20	3.75	28.41

AIC= -423.11

Figure G.5. Results of the tests of residuals for Model CS(100).

RUNS UP AND DOWN TEST		
Length	Observed	Expected
1	33	35.5000
2	16	15.3500
3 & up	5	5.4833

RUNS ABOVE AND BELOW THE MEAN TEST		
Length	Observed	Expected
1	20	21.7500
2	7	10.7500
3	4	5.31250
4 & up	8	5.18750

FREQUENCY TEST		
Interval	Observed	Expected
1	7	9.29050
2	9	9.46050
3	8	9.29900
4	9	9.71550
5	13	9.46900
6	13	9.71550
7	8	9.29900
8	9	9.46050
9	8	9.29050

PORTMANTEAU TEST		
dof	Q	Chi squared
1	0..01	2.71
2	0.02	4.61
3	0.05	6.25
4	0.08	7.78
5	0.09	8.24
6	0.15	10.64
7	0.17	12.02
8	0.21	13.36
9	0.52	14.68
10	0.58	15.99
11	0.66	17.27
12	0.72	18.55
13	0.83	19.81
14	0.89	21.06
15	1.20	22.31
16	1.30	23.54
17	1.49	24.77
18	1.83	25.99
19	1.93	27.20
20	2.07	28.41

AIC= -417.93

Figure G.6. Results of the tests of residuals for Model CW(100).

RUNS UP AND DOWN TEST

Length	Observed	Expected
1	30	35.5000
2	20	15.3500
3 & up	4	5.4833

RUNS ABOVE AND BELOW THE MEAN TEST

Length	Observed	Expected
1	21	21.7500
2	14	10.7500
3	5	5.31250
4 & up	4	5.18750

FREQUENCY TEST

Interval	Observed	Expected
1	8	9.29050
2	4	9.46050
3	12	9.29900
4	12	9.71550
5	14	9.46900
6	12	9.71550
7	9	9.29900
8	6	9.46050
9	7	9.29050

PORTMANTEAU TEST

dof	Q	Chi squared
1	0.06	2.71
2	0.35	4.61
3	0.62	6.25
4	1.14	7.78
5	1.49	8.24
6	1.78	10.64
7	2.07	12.02
8	2.35	13.36
9	2.62	14.68
10	3.12	15.99
11	3.43	17.27
12	3.75	18.55
13	5.62	19.81
14	6.57	21.06
15	7.44	22.31
16	8.03	23.54
17	8.97	24.77
18	9.69	25.99
19	10.73	27.20
20	11.25	28.41

AIC= -414.99

Figure G.7. Results of the tests of residuals for Model KN(100).

RUNS UP AND DOWN TEST		
Length	Observed	Expected
1	35	35.5000
2	19	15.3500
3 & up	3	5.4833

RUNS ABOVE AND BELOW THE MEAN TEST		
Length	Observed	Expected
1	22	21.7500
2	8	10.7500
3	9	5.31250
4 & up	3	5.18750

FREQUENCY TEST		
Interval	Observed	Expected
1	7	9.29050
2	7	9.46050
3	9	9.29900
4	12	9.71550
5	12	9.46900
6	12	9.71550
7	10	9.29900
8	7	9.46050
9	8	9.29050

PORTMANTEAU TEST		
dof	Q	Chi squared
1	0.03	2.71
2	0.05	4.61
3	0.07	6.25
4	0.20	7.78
5	0.62	8.24
6	0.73	10.64
7	0.34	12.02
8	0.94	13.36
9	1.11	14.68
10	1.23	15.99
11	1.41	17.27
12	1.57	18.55
13	1.75	19.81
14	1.93	21.06
15	2.15	22.31
16	2.28	23.54
17	2.64	24.77
18	2.83	25.99
19	2.99	27.20
20	3.39	28.41

AIC= -360.97

Figure G.8. Results of the tests of residuals for Model KE(100).

RUNS UP AND DOWN TEST		
Length	Observed	Expected
1	35	35.5000
2	15	15.3500
3 & up	6	5.4833

RUNS ABOVE AND BELOW THE MEAN TEST		
Length	Observed	Expected
1	30	21.7500
2	10	10.7500
3	5	5.31250
4 & up		5.18750

FREQUENCY TEST		
Interval	Observed	Expected
1	10	9.29050
2	6	9.46050
3	9	9.29900
4	15	9.71550
5	7	9.46900
6	9	9.71550
7	13	9.29900
8	4	9.46050
9	11	9.29050

PORTMANTEAU TEST		
dof	Q	Chi squared
1	0.03	2.71
2	0.04	4.61
3	0.06	6.25
4	0.19	7.78
5	0.38	8.24
6	0.47	10.64
7	0.57	12.02
8	0.65	13.36
9	0.73	14.68
10	0.81	15.99
11	0.95	17.27
12	1.24	18.55
13	1.92	19.81
14	2.57	21.06
15	2.71	22.31
16	2.91	23.54
17	3.16	24.77
18	3.34	25.99
19	3.53	27.20
20	3.75	28.41

AIC= -360.39

Figure G.9. Results of the tests of residuals for Model KS(100).

RUNS UP AND DOWN TEST		
Length	Observed	Expected
1	37	35.5000
2	18	15.3500
3 & up	3	5.4833

RUNS ABOVE AND BELOW THE MEAN TEST		
Length	Observed	Expected
1	13	21.7500
2	14	10.7500
3	8	5.31250
4 & up	4	5.18750

FREQUENCY TEST		
Interval	Observed	Expected
1	7	9.29050
2	10	9.46050
3	12	9.29900
4	13	9.71550
5	9	9.46900
6	10	9.71550
7	7	9.29900
8	7	9.46050
9	9	9.29050

PORTMANTEAU TEST		
dof	Q	Chi squared
1	0.06	2.71
2	0.10	4.61
3	0.17	6.25
4	0.27	7.78
5	0.33	8.24
6	0.49	10.64
7	0.58	12.02
8	0.65	13.36
9	0.72	14.68
10	0.80	15.99
11	0.88	17.27
12	0.98	18.55
13	1.67	19.81
14	1.95	21.06
15	2.14	22.31
16	2.52	23.54
17	2.72	24.77
18	3.16	25.99
19	3.40	27.20
20	3.57	28.41

AIC= -407.06

Figure G.10. Results of the tests of residuals for Model KW(100).

Bibliography

1. Allen, J.H. and J.D. Malick . "The Frequency of Cloud-Free Viewing Intervals." *AIAA 21st Aerospace Sciences Meeting.* Paper No. 83-0441. Reno Nevada: American Institute of Aeronautics and Astronautics, January 10–13, 1983.
2. Boehm, Albert R. and Donald D. Grantham. "The Effect of Viewing Aspect on Climatological Cloud Distribution." Proceedings of the International Conference on Statistical Climatology Abstracts (3rd). Vienna, Austria, 23-27 June 1986.
3. Box, George E. P. and Gwilym M. Jenkins. *Time Series Analysis Forcasting and Control.* San Francisco: Holden-Day, Inc., 1976.
4. Bradley, James V. *Distribution-Free Statistical Tests.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1968.
5. Brockwell, Peter J. and Richard A. Davis. *Time Series: Theory and Methods.* New York: Springer-Verlag, Inc., 1991.
6. Conover, W.J. *Practical Nonparametric Statistics.* New York: John Wiley & Sons, Inc., 1971.
7. Devore, Jay L. *Probability and Statistics for Engineering and the Sciences.* Pacific Grove, California: Brooks/Cole Publishing Company, 1991.
8. Feigelson, E.M. *Radiation in a Cloudy Atmosphere.* Boston: D. Reidel Publishing Company, 1984.
9. Grantham, Donald D. *Statistical Models For Cloud-Free Viewing Climatologies.* "Fourth International Conference on Statistical Climatology." Rotorua, New Zealand, 27-31 March 1989.
10. Grantham, Donald D., Chief, Atmospheric Structure Branch of the Atmospheric Sciences Division. Telephone interview. Phillips Laboratory, Geophysics Directorate, Hanscom Air Force Base, MA, 15 July 1992.
11. Hering, Wayne S. *Evaluation of Stochastic Models for Estimating the Persistence Probability of Cloud-Free Lines-of-Sight*, 1 October 1988–30 September 1989. Contract F19628-880K-0005. La Jolla, CA: Marine Physical Lab, Scripps Institution of Oceanography, September 1989.
12. Hering, Wayne S. *Probability Estimates of Cloud-Obscured Line-of-Sight.* Contract F19628-88-K-0005. La Jolla, CA: Marine Physical Lab, Scripps Institution of Oceanography, September 1989.
13. Johnson, R. W. and others. *Automated Visibility & Cloud Cover Measurements with a Solid-State Imaging System: Final Report*, 26 September 1984–25 September 1988. Contract F19628-84-K-0047. San Diego CA: Marine Physical Lab, Scripps Institution of Oceanography, March 1989.
14. Krishnaiah, P. R. and others. *Handbook of Statistics 5.* New York: Elsevier Science Publishing Company, Inc., 1985.
15. Lindgren, Bernard W. *Statistical Theory.* New York: Macmillan Publishing Co., Inc., 1976.
16. Lund, Iver A. and Milton D. Shanklin. "Photogrammetrically Determined Cloud-Free Lines-of-Sight Through the Atmosphere," *Journal of Applied Meteorology, 12:5* 773-782 (August 1972).
17. Lund, Iver A. and Milton D. Shanklin. "Universal Methods for Estimating Probabilities of Cloud-Free Lines-of-Sight Through the Atmosphere," *Journal of Applied Meteorology, 12:1* 28-35 (February 1973).

18. Lund, Iver A. "A Model for Estimating Joint Probabilities of Cloud-Free Lines-of-Sight Through the Atmosphere," it Journal of Applied Meteorology, 12:6 1040-1043 (September 1973).

19. Miller, Irwin R. and others. *Probability and Statistics for Engineers* (Fourth Edition). New Jersey: Prentice-Hall, Inc., 1990.

20. Milton, J.S. and Jesse C. Arnold. *Probability and Statistics in the Engineering and Computing Sciences.* New York: McGraw-Hill Book Company, 1986.

21. Naylor, Thomas H. and others. *Computer Simulation Techniques.* New York: John Wiley & Sons, 1991.

22. Powell, GEN Colin L., Chairman of the Joint Chiefs of Staff, "Dealing with the Changes." Address to the U.S. Naval Institute (Second Annapolis Seminar and 118th Annual Meeting). U.S. Naval Academy, Annapolis MD, 24 April 1992.

23. Schlotzhauer, Sandra D. and Ramon C. Littell. *SAS System for Elementary Statistical Analysis.* Cary, NC: SAS Institute, Inc., 1987.

24. Shields, Janet E. "Format of Ratio Tapes; One and Ten-Minute." Report (AV90-123t) to the Marine Physical Laboratory of the Scripps Institution of Oceanography. San Diego, CA, September 1990.

25. Shields, Janet E. "Format of One Minute Cloud Image Tapes." Report (AV89-082t) to the Marine Physical Laboratory of the Scripps Institution of Oceanography. San Diego, CA, August 1989.

26. Shields, Janet E., Senior Development Engineer for the Whole Sky Imager System. Telephone interview. Scripps Institution of Oceanography, San Diego CA, 15 October 1992.

27. STSC, Inc. *Statgraphics Reference Manual.* Rockville, Maryland: STSC, Inc., 1991.

28. Wolfe, William L. and George J. Zissis. *The Infrared Handbook.* Washington: Government Printing Office, 1978.

29. Yepez, Jeffrey, Physicist. Personal Correspondence. Phillips Laboratory, Geophysics Directorate, Hanscom Air Force Base, MA, 15 November 1991.

30. Yepez, Jeffrey, Physicist. Personal Correspondence. Phillips Laboratory, Geophysics Directorate, Hanscom Air Force Base, MA, 15 April 1992.

Vita

Captain Richard William Boltz was born on November 30, 1965. He graduated with honors from Marion Catholic High School in Marion, Ohio in June of 1984. In April of 1988, he graduated from the University of Dayton with a Bachelor of Science degree in Physics. He was a Reserve Officers Training Corps (ROTC) Distinguished Graduate and was awarded a regular commission upon graduation. In November of 1988, he married the former Cynthia Ann Lorei of Erie, Pennsylvania. In March of 1989, Captain Boltz was a Distinguished Graduate of Undergraduate Space Training, Lowry AFB, Colorado. After two years of operating Defense Support Program (DSP) and Defense Meteorological Satellite Program (DMSP) satellites in the Air Force's Mission Control Center 1-B, Falcon AFB, Colorado, Captain Boltz entered the School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.

Permanent address: 163 Barkswood Drive
Marion, Ohio 43302

Vita

Major George Richard Kather was born on November 10, 1957. He graduated with honors from Western High School in Shively, Kentucky in April of 1975. In June of 1979, he graduated from the United States Military Academy at West Point with a Bachelor of Science degree in General Engineering and was the 1979 recipient of the Abrams Award. He was commissioned as an Armor officer and was assigned to Fort Carson, Colorado where he served as a platoon leader, support platoon leader, and company executive officer in the 2nd Battalion of the 34th Armor. Upon reassignment to Gelnhausen, Federal Republic of Germany, he served as the Brigade Adjutant for 2nd Brigade of the 3rd Armor Division, and completed 32 months of command as Commander of B Company of 1-33 Armor, and Commander of HHC, 3-8 Cavalry. Following his tenure in command, Major Kather elected to single-track in his secondary specialty of Research and Development. He was assigned to the Tri-Service Standoff Attack Missile (TASSM) Program Office where he served as the Army's Assistant Program Manager for Logistics. He was selected as a member of the Army's Acquisition Corps in 1989 and graduated from the Defense Systems Management College's Program Manager's Course in September of 1990. Major Kather returned to the TSSAM Program Office and served as Chief of the Program Management Division until May of 1991 when he entered the School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.

Permanent address: 3005 Teakwood Circle
Shively, Kentucky 40216

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1992	Master's Thesis

4. TITLE AND SUBTITLE
ANALYSIS OF WHOLE-SKY IMAGER DATA TO DETERMINE THE VALIDITY OF PCFLOS MODELS

5. FUNDING NUMBERS

6. AUTHOR(S)
Richard W. Boltz and George R. Kather

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology, WPAFB OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/GEO/ENS/92D-04

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Phillips Laboratories
PL-GPAA
Hanscom AFB MA 01731-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
Models used to predict the probability of a cloud-free line-of-sight (PCFLOS) from the ground to space have existed since the 1960s. Unfortunately, an adequate data set has not been available to check the validity of these models until the deployment of the Whole-Sky Imager (WSI) system in 1989. Now that a three-year database has been collected from the WSI system, it is possible to validate, or refute, the existing models. This study investigates the most generally accepted models. Specifically, we investigate three questions: 1) Is the Lund and Shanklin PCFLOS model assumption of azimuthal independence valid; 2) Does the Lund and Shanklin sub-sampling of data via the use of a template adequately correlate to both the full image and the grid image; and, 3) Do the Lund and Shanklin and SRI model estimates correlate to the WSI observations. The primary contribution of this study is the development of a methodology which employs time series analysis techniques to evaluate and ultimately corroborate the assumption of azimuthal independence.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Whole-sky imager, PCFLOS, time series analysis	270
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL